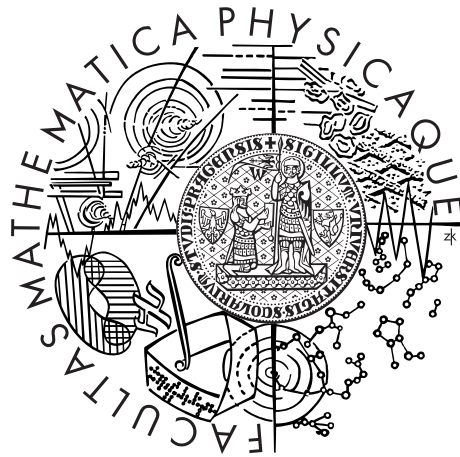


Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS



Marek Dvořák

Stability in Autoregressive Time Series Models

Department of Probability and Mathematical Statistics

Supervisor: Doc. RNDr. Zuzana Prášková, CSc.

Study programme: Econometrics and Oper. Research

Specialization: Econometrics

Prague 2015

First and foremost, I would like to express deepest thanks especially to my supervisor doc. RNDr. Zuzana Prášková, CSc. for her willingness to spend the valuable time for consultations and for reading many and many versions of my research articles and the contents of this thesis. Her careful guidance, mentorship and many useful comments significantly helped me to get over some difficulties during my research and offered me invaluable insight into the topic. My gratitude also goes to the group of researches around prof. RNDr. Marie Hušková, DrSc., with whom I had opportunity to speak at many conferences and mutual visits and which helped me to get in touch with the current progress in this research field. I strongly appreciate the support from RNDr. Martin Mádlík, Ph.D., who enabled me to use the computational cluster “Sněhurka”, where I could conduct most of the simulations. At the end, I appreciate a big support from the family, girlfriend and friends.

Acknowledgements

This work was supported by the grants

GAUK 586 712/2012,

GAČR 201/09/0755,

GAČR 201/09/J006,

GAČR P201/12/1277,

SVV 261 315/2011,

SVV 265 315/2012.

To my family.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, 30th September 2015

Marek Dvořák

Název práce: Stabilita v autoregresních modelech časových řad

Autor: Marek Dvořák

Katedra pravděpodobnosti a matematické statistiky

Vedoucí dizertační práce: Doc. RNDr. Zuzana Prášková, CSc.

Abstrakt: Předložená práce se zabývá oblastí detekce změn ve slabě stacionárních vektorových autoregresních modelech. Obsahem práce je návrh testových statistik pro retrospektivní detekci změny v různých parametrech těchto modelů a zejména odvození jejich asymptotického rozdělení za nulové hypotézy, kdy předpokládáme neměnnost těchto parametrů. Testové statistiky jsou založeny na principu maximální věrohodnosti a odvozeny za předpokladu normality, nicméně asymptotické výsledky u těchto statistik jsou platné pro daleko širší třídu rozdělení a zahrnují i modely, kde se vyskytují konkrétní formy závislosti. Součástí práce jsou rovněž simulační studie, které ilustrují kvalitu dosažených výsledků.

Klíčová slova: asymptotické rozdělení, change-point, slabá závislost, testování hypotéz, věrohodnost

Title: Stability in autoregressive time series models

Author: Marek Dvořák

Department: Department of Probability and Mathematical Statistics

Supervisor: Doc. RNDr. Zuzana Prášková, CSc.

Abstract: The main subject of this thesis is a change point detection in stationary vector autoregressions. Various test statistics are proposed for the retrospective break point detection in the parameters of such models, in particular, the derivation of their asymptotic distribution under the null hypothesis of no change. Testing procedures are based on the maximum likelihood principle and are derived under normality, nevertheless the asymptotic results are valid for broader class of distributions and involve also the models with certain form of dependence. Simulation studies document the quality of the results.

Keywords: asymptotic distribution, change-point, hypotheses testing, likelihood, weak dependence

Contents

1	Introduction	4
1.1	Testing procedures	4
1.1.1	Sequential monitoring	5
1.1.2	Retrospective tests	6
1.2	State of Art	6
2	VAR model, notation and assumptions	9
2.1	Model and hypothesis definition	9
2.2	Assumptions on the model	10
2.3	Proofs	15
3	LR test where variance of errors is unchanged	20
3.1	LR test derivation	20
3.2	Tests based on the approximation by functionals of Wiener process	22
3.3	Darling-Erdős-type test	24
3.4	Proofs	24
3.5	Simulation study	39
3.5.1	Known time point	40
3.5.2	Unknown time point	42
3.5.3	Change in the lag	43
3.5.4	Approximation by Gumbel distribution	43
3.5.5	Empirical power	44
3.6	Application	48
3.7	Chapter Summary	50
4	Darling-Erdős test for changes in variance	51
4.1	Test derivation	51
4.2	Darling-Erdős type approximation	52
4.3	Main result	56
4.4	Proofs	57
4.5	Simulation study	65
4.6	Chapter Summary	67
5	Score test	69
5.1	Introduction	69
5.2	Construction of the test statistic	70
5.3	Standardization matrix	71
5.4	Main result	73
5.5	Proofs	74
5.6	Simulation study	85
5.6.1	Simulations under the null hypothesis	85
5.6.2	Simulations under various alternatives	87
5.7	Chapter summary	92
6	Conclusion	93

A	Matrix algebra	95
B	Probability structures and asymptotic results	96
B.1	Martingale-differences	96
B.2	Stationarity and ergodicity	97
B.3	Strong mixings	98
B.4	M-dependent structures	99
	List of Abbreviations	101
	References	102

1. Introduction

Building time-series models with reasonable predictive strength plays essential role in Econometrics and Statistics. When developing such models, an analyst usually needs longer data to capture seasonal effects, to gain stationarity or to get valid statistical inference about the parameters. On the other hand, longer time series data are often exposed to sudden shocks such as financial crises, new regulations or changes in policy due to new political environment. These shocks may partly mask the key pieces of information behind data. One possibility is to build models with more parameters that can capture some of the effects however in this case they become too much complex and it is sometimes difficult to interpret such models to the practitioners. Another way to treat instability in the data is to find simpler models easier for interpretation and check whether there are time points (called *change-points*) where the parameters of the models become invalid and need to be changed. Some of the real data examples where the change points occur involve the quality control data, signal processing and segmentation, see for instance Basseville and Nikiforov (1993), temperature monitoring in e.g. Horváth et al. (2004), financial data in Bai (1997), or economic data in Horváth et al. (1997), among many others.

The popularity of the change point analysis has increased due to the vast amount of literature that covers various detectors of change points for many models of interest. For simple change in the location of normal random variables, see e.g. Yao and Davis (1986) or Horváth (1993b). For change points in regression models, see Horváth et al. (2004), Horváth et al. (2007) among others. Detecting changes in time series models are treated in Gombay (2008), Davis et al. (1995), Gombay and Horváth (2009), Hušková et al. (2007) and the references therein, or Prášková (2015) for change detection in random coefficient autoregression models. Multivariate time series models have been discussed for instance in Aue et al. (2009), Bai et al. (1998), or Bai (2000). Change point detection in panel data is treated in Chan et al. (2013), procedures for ARCH and GARCH sequences are discussed in Kokoszka and Leipus (2000), or Berkes et al. (2004). Testing procedures for finding breaks in functional observations are treated in Hörmann and Kokoszka (2010), among others.

The detection of a change usually involves two things that are closely related: First, a test has to be designed for the model of interest, where typically a null hypothesis means no change in the parameters of the model. Second, if a test rejects the null hypothesis, one should find the appropriate estimate of the change point. We will mainly discuss the tests since these are the primary interest of this thesis, and shortly mention also some references that deal with the properties of change point estimators.

1.1 Testing procedures

As regards the testing procedures, they can generally be split as in Figure 1.1.

The test statistics can be derived from residuals, by (quasi)-maximum likelihood approach such as test statistics based on score vector or likelihood ratio, or from Bayesian concept such as cumulative or moving sums (=CUSUM, MOSUM),

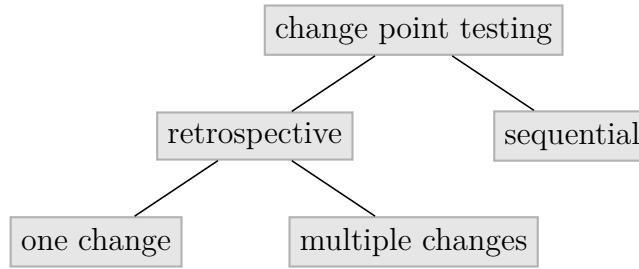


Figure 1.1: General split of change detectors.

etc. Test statistics are usually constructed in a way that large values indicate the rejection of the null hypothesis on a certain level and low values are in favour of the null. Since their exact distribution is sometimes impossible to find due to their complexity, the biggest effort is concentrated to find at least the asymptotic distribution under the null hypothesis and alternative. Such asymptotic distribution usually covers the distributions of several functions of Wiener processes, where the critical values have to be often simulated, or the exact distributions where we can use calculator to determine the critical value. Some Monte Carlo comparisons of various test statistics (Wald-type, likelihood-ratio, CUSUM, etc.) can be found in the paper by Andrews (1993) and the references therein.

1.1.1 Sequential monitoring

So called *sequential*, or *online* procedures are based on the real-time monitoring of the model. The procedure learns on some training sample where the estimation of the parameters is performed. Then the test statistic is updated every time the new observation comes, using the training sample estimates, and its value is compared to the critical value. If the statistic exceeds it we reject the null hypothesis and stop the monitoring.

The most popular in this context are CUSUM test statistics which were developed by Page (1955) in cases of shifts in the mean. One of the earliest well known application on ARMA models is the paper by Bagshaw and Johnson (1977) who found the asymptotic distribution of the test statistic under the null hypothesis based on the sums of squared residuals. As regards further papers on monitoring in time series models, we have to mention for instance the monitoring in autoregressions that is treated in Gombay and Serban (2009), or detecting changes in the variance structure for weakly stationary linear processes, see Gombay and Horváth (2009). Monitoring changes in linear models and asymptotic behaviour under alternative is treated for instance in Csörgő and Horváth (1997), Horváth et al. (2004), Hušková and Koubková (2005), Aue et al. (2006) where the latter paper also relaxes the independence condition in the error term. Weak dependence structure can also be found in Hušková and Chochola (2010), and L^p - m approximability conditions in Chochola et al. (2013). Darling-Erdős result in sequential setup is described for instance in Horváth et al. (2007).

1.1.2 Retrospective tests

In this setup analysts usually test some sample of a given length against changes that might occur. The conclusion of such analysis is a decision if there is (not) any substantial change point in such sample on a certain level of probability. The retrospective tests will be discussed in the thesis and the attention will be paid to the change point detection in autoregressions. We will consider tests based on the Gaussian quasi-likelihood approach.

As regards the historical development, one of the first articles that dealt with changes in the mean of Gaussian independent variables with unit variance was paper by Gardner (1969), where the time of change was assigned an apriori distribution. Likelihood approach for change point detection in the mean of continuous independent random variables is treated in Hinkley (1970) or Worsley (1986). Paper by Horváth (1993b) discusses the likelihood-ratio test for detecting changes in the mean and variance of independent normal observations and used the generalized the Darling-Erdős theorem, see Darling and Erdős (1956), to show that the standardized maximum of likelihood ratio follows Gumbel distribution under the null hypothesis of no change. Generalization for the latter case of changes in the mean where the errors in the location model follow stationary linear process is discussed in Horváth (1997). For the change detection in the regression models, see Horváth et al. (1997), or the more comprehensive book Csörgő and Horváth (1997). As regards the robust procedures for changes in regression, see for instance Hušková and Picek (2002) or Prášková and Chochola (2014).

One of the first papers considering structural changes in autoregressive models is the paper Salazar (1982) who investigates the change point in AR models by maximum likelihood method where the change point is assumed to follow certain distribution. Paper by Picard (1985) suggests tests in time series based on Kolmogorov-Smirnov idea for detecting failure in spectrum and for changes in parameters. Davis et al. (1995) develop Gaussian-type likelihood tests for detecting changes in stationary autoregressions. Hušková et al. (2007) suggests tests based on partial sums of weighted residuals and there is treated the behaviour of the test statistic also under alternative hypothesis as well. Changes in ARIMA models are covered in Lee et al. (2006).

1.2 State of Art

In this section we mention especially papers concerning testing changes in multivariate autoregression models and describe their contents in a bit more detail since these provides the baseline for us to propose some generalizations. Even if the articles often provide more results we will mention only the contributions that are directly linked with the topics covered in this thesis: Bai et al. (1998) considers the weak stationary vector autoregression model $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ with some additional stationary component and discusses the asymptotic behaviour of the F -type statistic. As regards the purely autoregression case, it is assumed that the error sequence is a martingale difference sequence with finite conditional variance; further, they assume finite $(4 + \delta)$ -moment, $\delta > 0$, of the error term. This implies that the error term must be uncorrelated with constant unconditional variance. In this case all the conditions apply on the error term. Article by Bai and Perron

(1998) discusses more general framework of multiple break points in linear models where lagged values of the process are (and are not) allowed as regressors. The assumptions are similar to those previously mentioned. Instead of higher moment conditions, it is assumed a certain FCLT to hold for the product of the error term and lagged value of the process. Controlling the empirical variance follows from the requirement that $1/T \cdot \sum_{t=1}^{T\tau} \mathbf{w}_t \mathbf{w}_t^\top \xrightarrow{P} \tau \mathbf{Q}$, uniformly in τ , where \mathbf{Q} is positive-definite and $\mathbf{w}_t = (\mathbf{y}_{t-1}^\top, \dots, \mathbf{y}_{t-p}^\top)^\top$ is a vector of lagged values of the process. This condition appears often in case of stochastic regressors, see Horváth (1995), Horváth et al. (2004), or Qu and Perron (2007) among many others where we usually do not have any specific information about their distribution. Qu and Perron (2007) consider quasi-likelihood ratio tests covering both multiple changes in the parameters of a general class of multivariate regression models and the case where prior information about the parameters is present. The versatility and universal usage on a variety of models is balanced off with some strict assumptions, some of which are difficult to verify.

We will further develop some of the testing procedures which were mentioned above. Chapter 2 contains basic notation and some sets of assumptions which will serve as a basis for the next chapters. The sets of assumptions are standard in the change point literature and contain weak dependence structures, see for instance Bai et al. (1998) for assumptions covering martingale-differences, Davis et al. (1995) for strongly mixing sequences, or Hörmann and Kokoszka (2010) for the L^p - m approximable sequences. All these weak dependent structures, as well as independent sequences, have in common that they imply FCLTs under further conditions on moments.

The main topic of Chapter 3 is a change detection in stationary VAR(p) models where the variance of the error term remains unchanged in time. The approximations of the quasi-likelihood test statistic by function of Wiener processes under the null has been studied in Qu and Perron (2007), however, we will not directly assume FCLTs. Instead, we will formulate the assumptions which will imply FCLT. In addition, a new VAR(p)-specific test, which can detect changes in the lag of the model, will be presented. We will also come up with Darling-Erdős test being inspired by Davis et al. (1995) where such tests are elaborated only for the univariate AR models. The part of this chapter has been published in Dvořák and Prášková (2013).

Chapter 4 is heavily based on the publication Dvořák (2015) which has recently been accepted to *Communication in Statistics – Theory and Methods*, and its content is devoted to the Darling-Erdős type test for situation when variance of the error term is also allowed to change. We will show that there is no direct generalization of the univariate case and explain why the test based on the classical quasi-log-likelihood ratio cannot converge to the Gumbel distribution under the null hypothesis, which is a quite surprising result since in the univariate case such approximation exists, see for instance Davis et al. (1995). We will propose the modification of the log-likelihood ratio in order to achieve the desired result under the null hypothesis. Some of the proofs are similar to Davis et al. (1995), however, lots of additional steps not treated in the latter article need to be proven to assure the convergence.

Chapter 5 presents the score test coming from the partial derivatives of quasi-likelihood ratio. The main idea comes from Gombay (2008) where the test is

introduced for change detection in univariate stationary autoregression. Unlike in previous two chapters, where we completely avoided independence, we will assume independent error process for deriving the asymptotic results. However, our assumptions will still be weaker than those in Gombay (2008) where a normality is assumed. We will show that the natural generalization to the multivariate set-up does not function in case of change detection in variance structure of the error term. Hence the new standardization matrix will be proposed specially for this test to be useful in the multivariate setting. Its advantage over the matrix used by Gombay (2008) stands in the fact that the test can be applied component-wise not only by blocks but also element by element.

So far, the estimates of the break points (break dates) have not been mentioned, since this is not a goal of the following text. Hence, we will give at least some relevant remarks and citations here. The estimate of the break point is usually an argument which maximizes the test statistic, see for instance Bai et al. (1998), p. 399, for the case of quasi-likelihood approach in vector autoregressions. The inference about the break points is extensively studied in Bai (2000). In particular it is shown under certain assumptions not stated here that in terms of real time index the break point estimate deviates from the true change point only by a finite number of observations. Asymptotic distribution for the break points estimators is also stated in that article. One of the most important result which links the estimates of the break dates and the estimators of the parameters in the quasi-likelihood approach is established in Qu and Perron (2007) in case of multiple breaks: They have found out that under certain set of conditions stated there, the maximization problem in the quasi-log-likelihood ratio over all possible change-points and all parameters can be divided into two asymptotically independent maximizations. First maximization problem finds the best estimators of the parameters with the change points being the true values, and, conversely, the second maximization problem finds the best estimators of the change points and does involve only true values of parameters. Hence, the estimation of the break points is not sensitive to the precision of the parameter estimators, and conclusively, the increasing length of the time series T does not help in the precision of the break dates estimators. According to this article, the precision of the break dates estimates can influence only changes in its true values across different regimes and the extent of the correlations in the error term.

2. VAR model, notation and assumptions

This chapter introduces the reader into the problem of change point detection in vector autoregressions. The first section contains the model definition and the considered scenarios of behaviour of the model in time. Section 2.2 discusses various sets of assumptions which have been studied in the literature and which will later be used to prove the desired properties of the test statistics. As it is difficult to compare the strength of these assumptions, we at least mention some important properties or implications for each of them. Section 2.3 contains proofs of the propositions stated in this chapter.

2.1 Model and hypothesis definition

Following usual definitions, see e.g. Hamilton (1994), we will denote by $\text{VAR}(p)$ a vector autoregressive model of the form

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \dots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad t \in \mathbb{Z}, \quad (2.1)$$

where p is a fixed lag, \mathbf{c} is an $n \times 1$ nonrandom vector, Φ_j , $j = 1, \dots, p$, are $n \times n$ nonrandom autoregressive matrices, and $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ is an n -dimensional error sequence that will be specified later.

In order to easier manipulate with the parameters we introduce the following notation: Let $\mathbf{V}_t = (1, \mathbf{y}_{t-1}^\top, \dots, \mathbf{y}_{t-p}^\top)^\top \in \mathbb{R}^{np+1}$, $\boldsymbol{\beta} = \text{vec}(\mathbf{c}, \Phi_1, \dots, \Phi_p) \in \mathbb{R}^{n(np+1)}$. Then we can write (2.1) in the form

$$\mathbf{y}_t = (\mathbf{V}_t^\top \otimes \mathbf{I}_n) \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t = \mathbf{M}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \quad t \in \mathbb{Z}, \quad (2.2)$$

where $\mathbf{M}_t := \mathbf{V}_t^\top \otimes \mathbf{I}_n \in \mathbb{R}^{n \times n(np+1)}$ and $\mathbf{I}_n := \mathbf{I}_{n \times n}$ stands for an n -dimensional identity matrix. Rewriting the model into the regression form (2.2) enables to express the test statistics in the more compact way.

Let us assume that we have T consecutive observations $\mathbf{y}_1, \dots, \mathbf{y}_T$ of process (2.2). Our aim is to test the null hypothesis

$$H_0 : \mathbf{y}_t = \mathbf{M}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \quad t = p+1, \dots, T,$$

against the following scenarios:

SCENARIO 1:

$$H_1 : \exists k \in \{p+1, \dots, T-1\} : \begin{aligned} \mathbf{y}_t &= \mathbf{M}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, & t = p+1, \dots, k, \\ \mathbf{y}_t &= \mathbf{M}_t \tilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}_t, & t = k+1, \dots, T, \end{aligned}$$

where $\tilde{\boldsymbol{\beta}} := \text{vec}(\tilde{\mathbf{c}}, \tilde{\Phi}_1, \dots, \tilde{\Phi}_p) \in \mathbb{R}^{n(np+1)}$, $\tilde{\boldsymbol{\beta}} \neq \boldsymbol{\beta}$.

In this scenario we assume that the variance of the error term is not subject to change however the variance of underlying process $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ can be subject of change since it depends on $\boldsymbol{\beta}$, see Hamilton (1994), p. 264, for the details in

stationary case.

SCENARIO 2:

$$H_1 : \exists k \in \{p+1, \dots, T-1\} : \begin{aligned} \mathbf{y}_t &= \mathbf{M}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, & t = p+1, \dots, k, \\ \mathbf{y}_t &= \widetilde{\mathbf{M}}_t \widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}_t, & t = k+1, \dots, T, \end{aligned} \quad (2.3)$$

where $\widetilde{\mathbf{M}}_t = (1, \mathbf{y}_{t-1}^\top, \dots, \mathbf{y}_{t-q}^\top)^\top \otimes \mathbf{I}_n \in \mathbb{R}^{n \times n(nq+1)}$, $\widetilde{\boldsymbol{\beta}} = \text{vec}(\widetilde{\mathbf{c}}, \widetilde{\boldsymbol{\Phi}}_1, \dots, \widetilde{\boldsymbol{\Phi}}_q) \in \mathbb{R}^{n(nq+1)}$, $p < q$.

The preceding SCENARIO 2 covers the case where autoregressive parameters can change at certain time k together with the increase of the lag of the model. Variance of the error term is supposed to be constant across the time.

SCENARIO 3:

$$H_1 : \exists k \in \{p+1, \dots, T-1\} : \begin{aligned} \mathbf{y}_t &= \mathbf{M}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, & t = p+1, \dots, k, \\ &= \mathbf{M}_t \widetilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}_t, & t = k+1, \dots, T, \end{aligned}$$

and either $\widetilde{\boldsymbol{\beta}} \neq \boldsymbol{\beta}$ or

$$\text{var}[\boldsymbol{\varepsilon}_t] = \begin{cases} \boldsymbol{\Omega}, & t = p+1, \dots, k \\ \widetilde{\boldsymbol{\Omega}}, & t = k+1, \dots, T \end{cases} \quad \text{where } \boldsymbol{\Omega} \neq \widetilde{\boldsymbol{\Omega}}.$$

The preceding scenario covers the case where both autoregression parameters and variance of the error term can be a subject of a change.

It will be assumed throughout the thesis in all scenarios that the initial p observations $\mathbf{y}_1, \dots, \mathbf{y}_p$ of the process $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ will follow VAR(p) process with parameter $\boldsymbol{\beta}$ and with variance of the error term $\boldsymbol{\Omega}$, to ensure that at least the first measurement of interest \mathbf{y}_{p+1} will obey the null hypothesis.

The time of change k is usually unknown, but can be estimated consistently by many test statistics, see for instance Bai et al. (1998), Bai (2000) for the case of change point in vector autoregression, or Qu and Perron (2007) in more general multivariate regression framework based on the quasi-likelihood approach. This thesis will deal only with testing structural changes.

2.2 Assumptions on the model

We will formulate various sets of assumptions which will further be used in the text. All assumptions have in common that they imply the functional central limit theorem which is the essential tool in the change point analysis for finding the asymptotic distribution of the test statistics.

ASSUMPTIONS A:

$$(A.1) \quad \forall |z| \leq 1 : \det \left\{ \mathbf{I}_n - \boldsymbol{\Phi}_1 z - \dots - \boldsymbol{\Phi}_p z^p \right\} \neq 0,$$

(A.2) Let $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ be an n -dimensional strictly stationary ergodic martingale difference sequence adapted to the filtration $\mathfrak{F}_t = \sigma\{\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \dots\}$ with $\mathbf{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top | \mathfrak{F}_{t-1}] = \boldsymbol{\Omega}$, where $\boldsymbol{\Omega} \in \mathbb{R}^{n \times n}$ is a known positive-definite variance matrix (which will be denoted as $\boldsymbol{\Omega} > 0$); further, let $\mathbf{E}[\varepsilon_{i,t} \varepsilon_{j,t} \varepsilon_{k,t} | \mathfrak{F}_{t-1}] = \mu_{i,j,k}$

2. VAR MODEL, NOTATION AND ASSUMPTIONS

be finite, $\forall t, \forall(i, j, k)$, and $\mathbf{E}[\varepsilon_{i,t} \varepsilon_{j,t} \varepsilon_{k,t} \varepsilon_{\ell,t} | \mathfrak{F}_{t-1}] = \mu_{i,j,k,\ell}$ be finite, $\forall t$, and $\forall(i, j, k, \ell)$.

Assumption (A.1) means that roots of the polynomial $\phi(z) := \det \{ \mathbf{I}_n - \Phi_1 z - \dots - \Phi_p z^p \}$ lie outside the complex unit circle. Assumption (A.2) yields FCLT for $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. Martingale difference property ensures that such sequence is uncorrelated with any measurable function of its lagged values, see for instance Davidson (1994), Theorem 15.3. Both assumptions imply that, under H_0 , the process as given in (2.2) can be represented as a vector infinite-order moving average process

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \varepsilon_{t-k}, \quad (2.4)$$

where $\boldsymbol{\mu} := \mathbf{E}[\mathbf{y}_t] = (\mathbf{I}_n - \sum_{j=1}^p \Phi_j)^{-1} \mathbf{c}$; and the sequence of matrices $\boldsymbol{\Psi}_k = \{\psi_{ij}^{(k)}\}_{i,j=1}^n$ is absolutely summable, i.e.

$$\sum_{k=0}^{\infty} |\psi_{ij}^{(k)}| < \infty, \quad \forall i, j = 1, \dots, n$$

(see, e.g. Hamilton (1994), Chapter 10.1., or Lütkepohl (2005), p. 657). It follows from Assumptions A and Theorem B.6, that under H_0 , the sequence $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ is strictly stationary and ergodic.

As we pointed out, it is well established in the literature that the elements $\psi_{ij}^{(k)}$ of the matrices $\boldsymbol{\Psi}_k$ in representation (2.4) are absolutely summable under Assumptions A. Other sets of assumptions presented later will also imply the infinite-order moving average representation. From the theory of univariate weakly stationary AR processes we know the speed of decay of the coefficients in such representation. Natural question therefore is, if something similar exists also in case of weakly stationary multivariate vector autoregressions. This question will be answered in Theorem 2.1. Now, we are going to prepare the setup for the proof of this theorem by transforming VAR(p) model into VAR(1) model:

Let us denote the following vectors

$$\mathbf{Y}_t := \text{vec}(\mathbf{y}_t, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p+1}), \quad \mathbf{C} := \text{vec}(\mathbf{c}, \mathbf{0}, \dots, \mathbf{0}), \quad \mathbf{U}_t := \text{vec}(\varepsilon_t, \mathbf{0}, \dots, \mathbf{0})$$

and matrices

$$\mathbf{A} := \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ \mathbf{I}_n & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_n & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{np \times np}, \quad \mathbf{J} := (\mathbf{I}_n \mathbf{0} \dots \mathbf{0}) \in \mathbb{R}^{n \times np}. \quad (2.5)$$

Process (2.1) can then be expressed as an (np) -dimensional VAR(1) process in the form

$$\mathbf{Y}_t = \mathbf{C} + \mathbf{A} \mathbf{Y}_{t-1} + \mathbf{U}_t,$$

which can be further written as

$$\mathbf{Y}_t = \boldsymbol{\nu} + \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{U}_{t-k}, \quad (2.6)$$

2. VAR MODEL, NOTATION AND ASSUMPTIONS

where $\boldsymbol{\nu} = \text{vec}(\boldsymbol{\mu}, \boldsymbol{\mu}, \dots, \boldsymbol{\mu}) \in \mathbb{R}^{np}$. Multiplying (2.6) by \mathbf{J} from the left and noticing that $\mathbf{U}_t = \mathbf{J}^\top \boldsymbol{\varepsilon}_t$ leads us to

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{k=0}^{\infty} \mathbf{J} \mathbf{A}^k \mathbf{J}^\top \boldsymbol{\varepsilon}_{t-k}. \quad (2.7)$$

We are about to show the rate of decay of the absolute value of elements $\psi_{ij}^{(k)}$. Throughout the text, let $\|\mathbf{A}\| := \sqrt{\text{tr}\{\mathbf{A}\mathbf{A}^\top\}}$, be the Euclidean norm of matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. It is proven that the summability of \mathbf{A} in the Euclidean norm is equivalent to the absolute summability of its elements, see for instance comment below Proposition C.8 in Lütkepohl (2005), p. 687-688.

Theorem 2.1 *Let $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ be a VAR(p) process defined in (2.1) fulfilling (A.1) with the error sequence $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ such that $\mathbb{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\varepsilon}_s \boldsymbol{\varepsilon}_t^\top] = \mathbb{I}_{[s=t]} \boldsymbol{\Omega} < \infty$ for all $s, t \in \mathbb{Z}$. Let \mathbf{A} be a matrix defined in (2.5). Then the process on the right-hand side of (2.4) is correctly specified and matrices $\boldsymbol{\Psi}_k$ in representation (2.4) fulfill inequality*

$$\|\boldsymbol{\Psi}_k\| \leq K \cdot k^{r_*-1} \cdot \lambda_*^k, \quad (2.8)$$

where $1 \leq K < \infty$ is a constant, $\lambda_* := \max_l \{|\lambda_l|\}$ is the largest eigenvalue of matrix \mathbf{A} in modulus, $0 < \lambda_* < 1$, and $r_* := \max_l \{r_l\}$, where r_l is a multiplicity of eigenvalue λ_l .

Conditions stated in Theorem 2.1 guarantee the existence of the infinite moving average representation of VAR(p) model in the mean square. The theorem tells us that the speed of decay of elements $\psi_{ij}^{(k)}$ of matrices $\boldsymbol{\Psi}_k$, $k = 0, 1, \dots$, depends on the solutions (eigenvalues) λ of the polynomial equation

$$\det\{\lambda \mathbf{I}_{np} - \mathbf{A}\} = \det\{\mathbf{I}_n \lambda^p - \boldsymbol{\Phi}_1 \lambda^{p-1} - \dots - \boldsymbol{\Phi}_{p-1} \lambda - \boldsymbol{\Phi}_p\} = 0. \quad (2.9)$$

Under stability conditions, if all eigenvalues of \mathbf{A} are mutually distinct then the elements $\psi_{ij}^{(k)}$, $i, j = 1, \dots, n$ decay geometrically when $k \rightarrow \infty$. Otherwise the rate of decay is somewhat slower proportional to k to the power of the biggest multiplicity of the root of equation (2.9).

Let us move to the other sets of conditions that will be used further:

ASSUMPTIONS B:

(B.1) $\forall |z| \leq 1 : \det\{\mathbf{I}_n - \boldsymbol{\Phi}_1 z - \dots - \boldsymbol{\Phi}_p z^p\} \neq 0.$

(B.2) Error term process $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ is a sequence of centered random vectors such that for all $t_1 \leq t_2 \leq t_3 \leq t_4$

a) $\mathbb{E}[\boldsymbol{\varepsilon}_{t_1} \boldsymbol{\varepsilon}_{t_2}^\top] = \mathbb{I}_{[t_1=t_2]} \cdot \boldsymbol{\Omega}$ is finite, and $\boldsymbol{\Omega} > 0$,

b) $\forall (i, j, k) : \mathbb{E}[\varepsilon_{i,t_1} \varepsilon_{j,t_2} \varepsilon_{k,t_3}] = \mathbb{I}_{[t_1=t_2=t_3]} \cdot \mu_{i,j,k}$ is finite,

c) $\forall(i, j, k, \ell)$:

$$\mathbf{E} \left[\varepsilon_{i,t_1} \varepsilon_{j,t_2} \varepsilon_{k,t_3} \varepsilon_{\ell,t_4} \right] = \begin{cases} \mu_{i,j,k,\ell} & \text{if } t_1 = t_2 = t_3 = t_4, \\ \sigma_{ij} \sigma_{kl} & \text{if } t_1 = t_2 < t_3 = t_4, \\ 0 & \text{otherwise,} \end{cases}$$

and $\mu_{i,j,k,\ell}$ is finite, for all i, j, k, ℓ , and σ_{ij} is (i, j) -element of $\mathbf{\Omega}$, $i, j = 1, \dots, n$.

d) $\sup_t \mathbf{E} \|\varepsilon_t\|^{4+\delta} = \text{const.} < \infty$, for some $\delta > 0$, i.e. process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ has uniformly bounded $(4 + \delta)$ -moment.

(B.3) Process $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ is a strong mixing process with rate $\rho_T = \mathcal{O}(T^{-(1+\epsilon)(1+4/\delta)})$ for some $\epsilon > 0$, where $\mathcal{O}(\cdot)$ denotes Landau symbol.

Assumption (B.1) is the same as (A.1) and guarantees stability of the VAR(p) model. Assumption (B.2) specifies the properties of the error term $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. Finally, (B.3) specifies the probabilistic structure of the observed process. According to Pham and Tran (1985) the mixing condition is satisfied for n -dimensional centered linear processes of the form $\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{\Psi}_j \varepsilon_{t-j}$, $\mathbf{\Psi}_0 = \mathbf{I}_n$, if

(Ph.1) ε_t are independent random vectors that admit density g_t such that $\forall t$ and $\forall \mathbf{u} \in \mathbb{R}^n$

$$\int_{\mathbb{R}^n} |g_t(\mathbf{v} - \mathbf{u}) - g_t(\mathbf{v})| d\mathbf{v} < K \|\mathbf{u}\| < \infty,$$

(Ph.2) it holds $\mathbf{E} \|\varepsilon_t\|^\delta < K$, $\forall t$, for some $\delta > 0$, $K > 0$,

(Ph.3) $\sum_{j=0}^{\infty} \|\mathbf{\Psi}_j\| < \infty$, $\sum_{j=1}^{\infty} \sum_{k=j}^{\infty} \|\mathbf{\Psi}_k\|^{\frac{\delta}{1+\delta}} < \infty$,

(Ph.4) $\sum_{j=0}^{\infty} \mathbf{\Psi}_j z^j \neq 0$, $\forall z$ such that $|z| \leq 1$.

Assuming that $\mathbf{E}[\mathbf{y}_t] = \mathbf{0}$, Condition (Ph.2) follows from Assumption (B.2). Condition (Ph.3) follows from Theorem 2.1 and from Lemma 2.4. Assumption (Ph.4) is satisfied under Assumption (B.1). Assumption (Ph.1) can be therefore seen as an additional assumption on the error term in order Assumption (B.3) is fulfilled.

Next we formulate another set of conditions being inspired by the recent papers Wu and Min (2005) and Aue et al. (2009):

ASSUMPTIONS C:

(C.1) $\forall |z| \leq 1 : \det \left\{ \mathbf{I}_n - \mathbf{\Phi}_1 z - \dots - \mathbf{\Phi}_p z^p \right\} \neq 0$.

(C.2) Let $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ be n -dimensional process such that

$$\varepsilon_t = \mathbf{f}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots), \quad t \in \mathbb{Z}, \quad (2.10)$$

where $\mathbf{f} : \mathbb{R}^{n' \times \infty} \rightarrow \mathbb{R}^n$ is a measurable function and $\{\boldsymbol{\nu}_t\}_{t \in \mathbb{Z}}$ a sequence of independent, identically distributed random vectors with values in $\mathbb{R}^{n'}$.

(C.3) It is further required that there is a sequence of m -dependent random vectors $\{\varepsilon_t^{(m)}\}_{t \in \mathbb{Z}}$ such that $\varepsilon_t^{(m)} = \mathbf{f}^{(m)}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots, \boldsymbol{\nu}_{t-m})$, $t \in \mathbb{Z}$, with measurable functions $\mathbf{f}^{(m)} : \mathbb{R}^{n' \times (m+1)} \rightarrow \mathbb{R}^n$, and

$$\sum_{m=1}^{\infty} \left(\mathbf{E} \|\varepsilon_0 - \varepsilon_0^{(m)}\|^4 \right)^{\frac{1}{4}} < \infty. \quad (2.11)$$

(C.4) $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ is a centered process, $\mathbb{E}[\boldsymbol{\varepsilon}_s \boldsymbol{\varepsilon}_t^\top] = \mathbb{I}_{[s=t]} \boldsymbol{\Omega}$, $\boldsymbol{\Omega} > 0$, with $\mathbb{E} \|\boldsymbol{\varepsilon}_t\|^4 < \infty$.

A disadvantage of Assumptions C is the need of certain structure on the error term sequence which is expressed in (2.10), whereas mixing assumptions can work with more general class of models. However, there is a lot of important models fulfilling (2.10), see Aue et al. (2009) for examples. In addition, mixing conditions are usually accompanied by additional smoothness restrictions which are often difficult to verify. Advantage is that Assumptions C can be applied even on possibly nonlinear error sequence.

The main difference of our approach and the paper by Aue et al. (2009) lies in the fact that we apply these assumptions on the error sequence while the authors of the latter paper use them for the time series itself. It has to be noted that the assumptions in Aue et al. (2009) do not contain (C.1) and (C.4). However, when the authors apply their theory on the case of VAR(p) process in Section 4.1, they assume the infinite-order moving average representation of VAR(p) process with iid error term sequence with bounded fourth absolute moment. Hence they implicitly assume (C.1) and (C.4) to be fulfilled as well. Unlike in Aue et al. (2009), we do not need the iid error term in our case. Their paper enables to widen the theory on ARCH and GARCH sequences. On the other side, they tackle only changes in variance of the series itself and they do not obtain which parameter causes these changes.

It follows from (C.2) and Theorem B.6 that error process $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ is strictly stationary and ergodic which is in line with Assumptions A. In (A.2) it is further required that the error sequence $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ possesses a martingale-difference property whereas here we assume uncorrelatedness in time and (C.3).

The concrete description of dependence, that is needed to establish the FCLTs, appears in (2.11). The weak dependence is enabled through the introduction of m -dependent random errors $\{\boldsymbol{\varepsilon}_t^{(m)}\}_{t \in \mathbb{Z}}$ in condition (C.3). It does not mean that the original error sequence $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ is m -dependent, it must only be “close” to $\boldsymbol{\varepsilon}_t^{(m)}$ in the sense of (2.11). Paper by Aue et al. (2009) shows that (C.2) and (2.11), where $\boldsymbol{\varepsilon}_t$ is replaced by \mathbf{y}_t , induce, under H_0 , FCLT for standardized \mathbf{y}_t and $\text{vech}(\mathbf{y}_t \mathbf{y}_t^\top)$, see Theorem B.11 and Theorem B.12.

There has been developed the theory of L^p - m approximable sequences where the main effort stands in avoiding the mixing conditions, see Wu and Min (2005) for the case of linear processes, or Hörmann and Kokoszka (2010) for the case of weakly dependent functional data. As pointed out in Hörmann and Kokoszka (2010), the L^p - m approximable sequences do not imply the strong mixing conditions and that the concept of L^p - m approximability is not directly comparable to the mixing processes.

Now we will show that if $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ follows Assumptions C then the stationary VAR(p) model $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ follows Assumption 2.1 in Aue et al. (2009) and hence FCLT can be applied to \mathbf{y}_t and $\text{vech}(\mathbf{y}_t \mathbf{y}_t^\top)$ under H_0 , see Theorems B.11 and B.12.

Theorem 2.2 *Let Assumptions C be fulfilled. Then there exists a measurable function $\mathbf{g} : \mathbb{R}^{n' \times \infty} \rightarrow \mathbb{R}^n$ such that*

$$\mathbf{y}_t = \mathbf{g}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots), \quad t \in \mathbb{Z}, \quad (2.12)$$

where $\{\boldsymbol{\nu}_t\}_{t \in \mathbb{Z}}$ is a sequence of iid random vectors with values in $\mathbb{R}^{n'}$. Further, there exists $m > 0$ and a sequence of $(2m)$ -dependent random vectors $\{\mathbf{y}_t^{(2m)}\}_{t \in \mathbb{Z}}$

such that $\mathbf{y}_t^{(2m)} = \mathbf{g}^{(2m)}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots, \boldsymbol{\nu}_{t-2m})$, $t \in \mathbb{Z}$, with measurable functions $\mathbf{g}^{(2m)} : \mathbb{R}^{n' \times (2m+1)} \rightarrow \mathbb{R}^n$, and

$$\sum_{m=1}^{\infty} \left(\mathbb{E} \left\| \mathbf{y}_0 - \mathbf{y}_0^{(2m)} \right\|^4 \right)^{\frac{1}{4}} < \infty. \quad (2.13)$$

In particular, Assumptions A and B imply FCLTs as well, as will become clear further. Since the convergence in FCLT leads to functions of the Wiener process, let us remind its definition:

Definition 2.3 We say, that \mathbf{W}_Γ is a d -dimensional Wiener process with covariance matrix Γ , if it is a centered Gaussian process with covariance function $\text{Cov}(\mathbf{W}_\Gamma(s), \mathbf{W}_\Gamma(t)) = \min\{s, t\}\Gamma$. We say that \mathbf{W} is a d -dimensional standard Wiener process, if $\Gamma = \mathbf{I}_d$. Process \mathbf{B} is a d -dimensional standard Brownian bridge, if $\mathbf{B}(\tau) = \mathbf{W}(\tau) - \tau\mathbf{W}(1)$, $0 \leq \tau \leq 1$, where \mathbf{W} is a d -dimensional standard Wiener process.

2.3 Proofs

In this section the statements of Chapter 2 will be proven.

Proof of Theorem 2.1: Comparing the equation (2.7) with (2.4) we see that $\boldsymbol{\Psi}_k = \mathbf{J}\mathbf{A}^k\mathbf{J}^\top$. The Jordan decomposition of the matrix \mathbf{A} is of the form $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$, where \mathbf{P} is regular and $\boldsymbol{\Lambda}$ is a block-diagonal Jordan matrix of the form

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Lambda}_s \end{pmatrix}, \quad \boldsymbol{\Lambda}_l = \begin{pmatrix} \lambda_l & 1 & 0 & \dots & 0 \\ 0 & \lambda_l & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \dots & \lambda_l \end{pmatrix} \in \mathbb{C}^{r_l \times r_l},$$

where $\lambda_1, \dots, \lambda_s$ are mutually distinct eigenvalues of \mathbf{A} less than 1 in modulus, $s \leq np$, and r_l is a multiplicity of eigenvalue λ_l , $l = 1, \dots, s$. Note that if multiplicity r_l of eigenvalue λ_l is 1, then the Jordan block $\boldsymbol{\Lambda}_l$ is a number, i.e. $\boldsymbol{\Lambda}_l = \lambda_l$. By multiplication we get that $\mathbf{A}^k = \mathbf{P}\boldsymbol{\Lambda}^k\mathbf{P}^{-1}$, where

$$\boldsymbol{\Lambda}^k = \begin{pmatrix} \boldsymbol{\Lambda}_1^k & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2^k & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Lambda}_s^k \end{pmatrix}, \quad \boldsymbol{\Lambda}_l^k = \begin{pmatrix} \lambda_l^k & \binom{k}{1}\lambda_l^{k-1} & \dots & \binom{k}{r_l-1}\lambda_l^{k-r_l+1} \\ 0 & \lambda_l^k & \dots & \binom{k}{r_l-2}\lambda_l^{k-r_l+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_l^k \end{pmatrix},$$

$l = 1, \dots, s$, $k = 0, 1, \dots$, and binomial coefficient is defined as

$$\binom{a}{b} := \begin{cases} \frac{a!}{b!(a-b)!} & \text{for } a \geq b, \\ 0 & \text{for } a < b, \end{cases}$$

with $0! := 1$. The norm of the moving average terms is of the form

$$\|\boldsymbol{\Psi}_k\| = \|\mathbf{J}\mathbf{P}\boldsymbol{\Lambda}^k\mathbf{P}^{-1}\mathbf{J}^\top\| \leq \|\mathbf{J}\| \cdot \|\mathbf{P}\| \cdot \|\boldsymbol{\Lambda}^k\| \cdot \|\mathbf{P}^{-1}\| \cdot \|\mathbf{J}^\top\| \leq K\|\boldsymbol{\Lambda}^k\|,$$

$k = 0, 1, \dots$, with positive finite constant K .

For the rest of the proof we will focus on the norm of the block-diagonal matrix $\mathbf{\Lambda}^k$. By direct computation

$$\begin{aligned}
 \|\mathbf{\Lambda}^k\|^2 &= \text{tr}\left\{\mathbf{\Lambda}^k(\mathbf{\Lambda}^k)^\top\right\} = \\
 &= \sum_{l=1}^s \left(\lambda_l^{2k} + \left[\binom{k}{1}\lambda_l^{k-1}\right]^2 + \dots + \left[\binom{k}{r_l-1}\lambda_l^{k-r_l+1}\right]^2 + \right. \\
 &\quad \left. + \lambda_l^{2k} + \left[\binom{k}{1}\lambda_l^{k-1}\right]^2 + \dots + \left[\binom{k}{r_l-2}\lambda_l^{k-r_l+2}\right]^2 + \dots \right. \\
 &\quad \left. \dots + \lambda_l^{2k} \right) = \\
 &= \sum_{l=1}^s \left(\tilde{\lambda}_l^k + \left[\binom{k}{1}\right]^2 \tilde{\lambda}_l^{k-1} + \dots + \left[\binom{k}{r_l-1}\right]^2 \tilde{\lambda}_l^{k-r_l+1} + \right. \\
 &\quad \left. + \tilde{\lambda}_l^k + \left[\binom{k}{1}\right]^2 \tilde{\lambda}_l^{k-1} + \dots + \left[\binom{k}{r_l-2}\right]^2 \tilde{\lambda}_l^{k-r_l+2} + \dots \right. \\
 &\quad \left. \dots + \tilde{\lambda}_l^k \right) =: \sum_{l=1}^s z_l,
 \end{aligned}$$

where $\tilde{\lambda}_l := \lambda_l^2$, $0 < \tilde{\lambda}_l < 1$, $l = 1, \dots, s$.

If $r_l = 1$ then $z_l = \tilde{\lambda}_l^k$. Otherwise, if $r_l > 1$ then in this case

$$z_l = \sum_{u=0}^{r_l-1} \left((r_l - u) \left[\binom{k}{u}\right]^2 \tilde{\lambda}_l^{k-u} \right).$$

Let us split the set $S := \{1, \dots, s\}$ on two parts S_1, S_2 where $S_1 = \{l \in S : r_l = 1\}$ and $S_2 = \{l \in S : r_l > 1\}$. It is clear that $S_1 \cap S_2 = \emptyset$ and $S_1 \cup S_2 = S$. Then

$$\|\mathbf{\Lambda}^k\|^2 = \underbrace{\sum_{l \in S_1} \tilde{\lambda}_l^k}_{\text{Term (A)}} + \underbrace{\sum_{l \in S_2} \sum_{u=0}^{r_l-1} \left((r_l - u) \left[\binom{k}{u}\right]^2 \tilde{\lambda}_l^{k-u} \right)}_{\text{Term (B)}}. \quad (2.14)$$

Further, let K be a positive finite constant which can change throughout the rest of the proof from line to line. Let us consider two options:

Situation (1):

If $r_l = 1$ for all $l = 1, \dots, s$, then $s = np$, $S_2 = \emptyset$ and \mathbf{A} has (np) -distinct eigenvalues. In that case only Term (A) in the equality (2.14) applies and elements of $\mathbf{\Psi}_k$ decay geometrically, since $\|\mathbf{\Lambda}^k\| \leq \sqrt{np} \cdot \lambda_*^k$, where $\lambda_* := \max_{l=1, \dots, s} \{|\lambda_l|\}$.

Situation (2):

Otherwise, there exists $l \in \{1, \dots, s\}$ such that $r_l > 1$. In this case $S_2 \neq \emptyset$ and Term (B) in (2.14) has to be taken into consideration.

2. VAR MODEL, NOTATION AND ASSUMPTIONS

Case (2a):

If $k \geq 2(r_l - 1)$, then there exists $K > 0$ such that

$$z_l \leq r_l^2 \left[\binom{k}{r_l - 1} \right]^2 \cdot \tilde{\lambda}_l^{k - r_l + 1} \leq K \cdot k^{2(r_l - 1)} \cdot \lambda_l^{2k}. \quad (2.15)$$

Case (2b):

If $1 \leq k < 2(r_l - 1)$, then the maximal summand in Term (B) lies somewhere “in-between” (r_l) -combinatorial terms, i.e. there exists τ_l depending on k, r_l, λ_l such that

$$z_l \leq r_l(r_l - \tau_l) \left[\binom{k}{\tau_l} \right]^2 \cdot \tilde{\lambda}_l^{k - \tau_l} \leq K \cdot k^{2\tau_l} \cdot \lambda_l^{2k}.$$

Since, in case (2b), $1 \leq \tau_l < r_l - 1$, then

$$z_l \leq K \cdot k^{2(r_l - 1)} \cdot \lambda_l^{2k}. \quad (2.16)$$

If we denote $r_* := \max_{l=1, \dots, s} \{r_l\}$ then, combining (2.14), (2.15) and (2.16), the upper-bound for $\|\mathbf{\Lambda}^k\|^2$ becomes

$$\begin{aligned} \|\mathbf{\Lambda}^k\|^2 &\leq \sum_{l \in S_1} \lambda_l^{2k} + \sum_{l \in S_2} \left(K \cdot k^{2(r_l - 1)} \cdot \lambda_l^{2k} \right) \leq \\ &\leq K \cdot k^{2(r_* - 1)} \sum_{l \in S} \lambda_l^{2k} \leq K \cdot k^{2(r_* - 1)} \cdot \lambda_*^{2k}, \end{aligned}$$

$1 \leq K < \infty$, and from that we obtain by taking the square root

$$\|\mathbf{\Lambda}^k\| \leq K \cdot k^{r_* - 1} \cdot \lambda_*^k$$

which is the assertion. □

Lemma 2.4 *Let $r \in \mathbb{N}_0$ and $0 < \lambda < 1$. Then the double-infinite series*

$$\sum_{j=1}^{\infty} \sum_{k=j}^{\infty} k^r \lambda^k$$

converges.

Proof: The result is immediate after re-arranging the summations,

$$\sum_{j=1}^{\infty} \sum_{k=j}^{\infty} k^r \lambda^k = \sum_{k=0}^{\infty} k^{r+1} \lambda^k.$$

□

Before proving Theorem 2.2, we will need an auxiliary lemma which is a Minkowski’s inequality for random vectors. Let us denote for $p \geq 1$, and random vector $\mathbf{x} \in \mathbb{R}^d$ with $\mathbb{E} \|\mathbf{x}\|^p < \infty$, where $\|\mathbf{x}\| = (\sum_{i=1}^d x_i^2)^{\frac{1}{2}}$,

$$\nu_p(\mathbf{x}) := \left(\mathbb{E} \left[\|\mathbf{x}\|^p \right] \right)^{\frac{1}{p}}.$$

This is a generalization of univariate \mathcal{L}^p norm. We will now show that Minkowski’s inequality holds and therefore that $\nu_p(\cdot)$ is a norm in d -dimensional \mathcal{L}^p space:

2. VAR MODEL, NOTATION AND ASSUMPTIONS

Lemma 2.5 For $p \geq 1$, and for random vectors \mathbf{x} and \mathbf{y} such that $\nu_p(\mathbf{x}) < \infty$ and $\nu_p(\mathbf{y}) < \infty$ it holds

$$\nu_p(\mathbf{x} + \mathbf{y}) \leq \nu_p(\mathbf{x}) + \nu_p(\mathbf{y}). \quad (2.17)$$

Proof: Let us recall that by properties of Euclidean norm and Hölder's inequality

$$\begin{aligned} (\nu_p(\mathbf{x} + \mathbf{y}))^p &= \mathbf{E} \left[\|\mathbf{x} + \mathbf{y}\| \cdot \|\mathbf{x} + \mathbf{y}\|^{p-1} \right] \leq \\ &\leq \mathbf{E} \left[\|\mathbf{x}\| \cdot \|\mathbf{x} + \mathbf{y}\|^{p-1} \right] + \mathbf{E} \left[\|\mathbf{y}\| \cdot \|\mathbf{x} + \mathbf{y}\|^{p-1} \right] \leq \\ &\leq \left(\left(\mathbf{E} \|\mathbf{x}\|^p \right)^{\frac{1}{p}} + \left(\mathbf{E} \|\mathbf{y}\|^p \right)^{\frac{1}{p}} \right) \cdot \left(\mathbf{E} \left[\|\mathbf{x} + \mathbf{y}\|^p \right] \right)^{1 - \frac{1}{p}} = \\ &= \left(\nu_p(\mathbf{x}) + \nu_p(\mathbf{y}) \right) \cdot \frac{(\nu_p(\mathbf{x} + \mathbf{y}))^p}{\nu_p(\mathbf{x} + \mathbf{y})}. \end{aligned}$$

The conclusion is immediate by multiplying both sides of inequality by positive term $\frac{\nu_p(\mathbf{x} + \mathbf{y})}{(\nu_p(\mathbf{x} + \mathbf{y}))^p}$. \square

Proof of Theorem 2.2: It follows from Assumptions C that

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{t-j} = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{f}(\boldsymbol{\nu}_{t-j}, \boldsymbol{\nu}_{t-1-j}, \dots) =: \mathbf{g}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots),$$

where \mathbf{g} is measurable and hence (2.12) is fulfilled for \mathbf{y}_t .

Since $\boldsymbol{\varepsilon}_t^{(m)} = \mathbf{f}^{(m)}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots, \boldsymbol{\nu}_{t-m})$, $t \in \mathbb{Z}$, are m -dependent for some m , we can define $(2m)$ -dependent vectors

$$\begin{aligned} \mathbf{y}_t^{(2m)} &:= \boldsymbol{\mu} + \sum_{j=0}^m \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{t-j}^{(m)} = \boldsymbol{\mu} + \sum_{j=0}^m \boldsymbol{\Psi}_j \mathbf{f}^{(m)}(\boldsymbol{\nu}_{t-j}, \dots, \boldsymbol{\nu}_{t-j-m}) =: \\ &=: \mathbf{g}^{(2m)}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots, \boldsymbol{\nu}_{t-2m}). \end{aligned} \quad (2.18)$$

Let us check (2.13): In the following, K_1 and K_2 are positive finite constants, r is the biggest multiplicity of the roots of characteristic polynomial of $\text{VAR}(p)$ model as in Theorem 2.1. By applying Minkowski's inequality with $p = 4$ in Lemma 2.5

$$\begin{aligned} &\sum_{m=1}^{\infty} \left(\mathbf{E} \left\| \mathbf{y}_0 - \mathbf{y}_0^{(2m)} \right\|^4 \right)^{\frac{1}{4}} = \sum_{m=1}^{\infty} \nu_4(\mathbf{y}_0 - \mathbf{y}_0^{(2m)}) = \\ &= \sum_{m=1}^{\infty} \nu_4 \left(\sum_{j=0}^m \boldsymbol{\Psi}_j (\boldsymbol{\varepsilon}_{-j} - \boldsymbol{\varepsilon}_{-j}^{(m)}) + \sum_{j=m+1}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{-j} \right) \leq \\ &\leq \sum_{m=1}^{\infty} \nu_4 \left(\sum_{j=0}^m \boldsymbol{\Psi}_j (\boldsymbol{\varepsilon}_{-j} - \boldsymbol{\varepsilon}_{-j}^{(m)}) \right) + \sum_{m=1}^{\infty} \nu_4 \left(\sum_{j=m+1}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{-j} \right) \leq \\ &\leq \sum_{m=1}^{\infty} \sum_{j=0}^m \left(\|\boldsymbol{\Psi}_j\| \cdot \nu_4(\boldsymbol{\varepsilon}_0 - \boldsymbol{\varepsilon}_0^{(m)}) \right) + \sum_{m=1}^{\infty} \sum_{j=m+1}^{\infty} \left(\|\boldsymbol{\Psi}_j\| \cdot \nu_4(\boldsymbol{\varepsilon}_0) \right) \leq \\ &\leq \underbrace{K_1 \sum_{j=0}^{\infty} (j^{r_*-1} \lambda_*^j)}_{\text{Term (a)}} \cdot \underbrace{\sum_{m=1}^{\infty} \left(\mathbf{E} \left\| \boldsymbol{\varepsilon}_0 - \boldsymbol{\varepsilon}_0^{(m)} \right\|^4 \right)^{\frac{1}{4}}}_{\text{Term (b)}} + \underbrace{K_2 \sum_{m=1}^{\infty} \sum_{j=m+1}^{\infty} (j^{r_*-1} \lambda_*^j)}_{\text{Term (c)}}, \end{aligned}$$

2. VAR MODEL, NOTATION AND ASSUMPTIONS

where the last inequality follows from Theorem 2.1 with $0 < \lambda_* < 1$, $r_* \in \mathbb{N}$ specified in that theorem as well. Then Term (a) is a convergent numeric series, Term (b) converges according to (2.11). Finally, Term (c) is convergent due to Lemma 2.4. \square

3. LR test where variance of errors is unchanged

In the following two chapters, our approach will focus on change detection in VAR(p) models based on the likelihood ratio principle. We will assume throughout this chapter that unconditional variance Ω of the error terms remains constant across the sample, whereas in Chapter 4 we will allow also for changes in Ω .

The chapter is organized as follows: In Section 3.1 the quasi-likelihood ratio of VAR(p) process is derived. Section 3.2 contains the theorem about the approximation of the test statistic under H_0 by a function of Gaussian random process when a time of change is a priori known. The result is generalized afterwards for testing a change in a time interval and for the case we permit also a change in the lag of the model. The results are shown under various forms of assumptions which were introduced in the previous chapter. Section 3.3 discusses the Darling-Erdős type test which shows that the asymptotic distribution of the likelihood ratio-type test statistic properly normalized is a Gumbel distribution under H_0 . The usage of Gumbel distribution is new in the context of multivariate models and the derivation of the asymptotic distribution of the test statistic is based on the extension of idea elaborated in Davis et al. (1995) for univariate AR models. It utilizes a multivariate extension of the Darling-Erdős extremal theorem established by Horváth (1993b) and is heavily based on the article Dvořák and Prášková (2013). Auxiliary lemmas and proofs of the theorems are presented in Section 3.4. Several simulation examples are shown in Section 3.5 and they support the results of the theory. The simulation concept is quite wide and contains examples of stationary VAR(p) models which both follow and contradict the assumptions presented in the theoretical part. Since the theoretical parts do not discuss the behaviour of the test statistic under the alternative hypothesis we want to fill this gap at least in the simulation part. Section 3.6 contains a real data application example.

3.1 LR test derivation

Under the assumption that Ω is positive-definite, i.e. $\Omega > 0$, and unchanged we utilize the hypotheses in SCENARIO 1 to the form:

$$\begin{aligned} H_0 : \quad \Omega^{-\frac{1}{2}}\mathbf{y}_t &= \Omega^{-\frac{1}{2}}\mathbf{M}_t\boldsymbol{\beta} + \Omega^{-\frac{1}{2}}\boldsymbol{\varepsilon}_t, & t = p + 1, \dots, T, \\ H_1 : \quad \exists k : \quad \Omega^{-\frac{1}{2}}\mathbf{y}_t &= \Omega^{-\frac{1}{2}}\mathbf{M}_t\boldsymbol{\beta} + \Omega^{-\frac{1}{2}}\boldsymbol{\varepsilon}_t, & t = p + 1, \dots, k, \\ & \Omega^{-\frac{1}{2}}\mathbf{y}_t &= \Omega^{-\frac{1}{2}}\mathbf{M}_t\tilde{\boldsymbol{\beta}} + \Omega^{-\frac{1}{2}}\boldsymbol{\varepsilon}_t, & t = k + 1, \dots, T. \end{aligned}$$

Let us define $\mathbf{y}_t^* := \Omega^{-\frac{1}{2}}\mathbf{y}_t$, $\mathbf{M}_t^* := \Omega^{-\frac{1}{2}}\mathbf{M}_t$ and $\boldsymbol{\varepsilon}_t^* := \Omega^{-\frac{1}{2}}\boldsymbol{\varepsilon}_t$. Under Assumptions A, or B, or C, it holds that $\mathbb{E}[\boldsymbol{\varepsilon}_t^*] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\varepsilon}_s^* \boldsymbol{\varepsilon}_t^{*\top}] = \mathbf{I}_n \cdot \mathbb{I}_{[s=t]}$.

To derive the LR test we will assume the quasi-likelihood function (i.e. it is generally based on the Gaussian likelihood of the error term $\boldsymbol{\varepsilon}_t$). It is a common practice to use Gaussian inference tools in time series analysis even if the process is not a Gaussian-type.

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

Let us assume that $k \in \{p + 1, \dots, T - 1\}$ is fixed. Then the Gaussian likelihood ratio, conditional on the first p observations, is given by

$$LR = \frac{\prod_{t=p+1}^k f_t(\boldsymbol{\beta}) \cdot \prod_{t=k+1}^T f_t(\tilde{\boldsymbol{\beta}})}{\prod_{t=p+1}^T f_t(\boldsymbol{\beta})}, \quad (3.1)$$

where for $\mathbf{b} = \boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}$, respectively,

$$f_t(\mathbf{b}) = (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t^* - \mathbf{M}_t^* \mathbf{b})^\top (\mathbf{y}_t^* - \mathbf{M}_t^* \mathbf{b}) \right\}.$$

After taking log transformation, we obtain the test statistic based on LR of the form

$$\begin{aligned} \Lambda_T(k) &:= \min_{\boldsymbol{\beta}} \left\{ \sum_{t=p+1}^T (\mathbf{y}_t^* - \mathbf{M}_t^* \boldsymbol{\beta})^\top (\mathbf{y}_t^* - \mathbf{M}_t^* \boldsymbol{\beta}) \right\} \\ &\quad - \min_{\boldsymbol{\beta}} \left\{ \sum_{t=p+1}^k (\mathbf{y}_t^* - \mathbf{M}_t^* \boldsymbol{\beta})^\top (\mathbf{y}_t^* - \mathbf{M}_t^* \boldsymbol{\beta}) \right\} \\ &\quad - \min_{\tilde{\boldsymbol{\beta}}} \left\{ \sum_{t=k+1}^T (\mathbf{y}_t^* - \mathbf{M}_t^* \tilde{\boldsymbol{\beta}})^\top (\mathbf{y}_t^* - \mathbf{M}_t^* \tilde{\boldsymbol{\beta}}) \right\}. \end{aligned}$$

If we carry out the minimization and insert the respective arguments (i.e., the corresponding least squares estimators) $\hat{\boldsymbol{\beta}}_T$, $\hat{\boldsymbol{\beta}}_k$ and $\tilde{\hat{\boldsymbol{\beta}}}_k$ back into $\Lambda_T(k)$ we obtain

$$\begin{aligned} \Lambda_T(k) &= \sum_{t=p+1}^T (\mathbf{y}_t - \mathbf{M}_t \hat{\boldsymbol{\beta}}_T)^\top \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \mathbf{M}_t \hat{\boldsymbol{\beta}}_T) - \\ &\quad - \sum_{t=p+1}^k (\mathbf{y}_t - \mathbf{M}_t \hat{\boldsymbol{\beta}}_k)^\top \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \mathbf{M}_t \hat{\boldsymbol{\beta}}_k) \\ &\quad - \sum_{t=k+1}^T (\mathbf{y}_t - \mathbf{M}_t \tilde{\hat{\boldsymbol{\beta}}}_k)^\top \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \mathbf{M}_t \tilde{\hat{\boldsymbol{\beta}}}_k), \end{aligned} \quad (3.2)$$

which can be further written in a form more suitable for theoretical considerations as

$$\Lambda_T(k) = -\mathbf{s}_T^\top \mathbf{P}_T \mathbf{s}_T + \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k + \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k, \quad (3.3)$$

where for $k = p + 1, \dots, T - 1$,

$$\mathbf{s}_k = \sum_{t=p+1}^k \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t, \quad \tilde{\mathbf{s}}_k = \sum_{t=k+1}^T \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t, \quad (3.4)$$

$$\mathbf{P}_k = \left(\sum_{t=p+1}^k \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t \right)^{-1}, \quad \tilde{\mathbf{P}}_k = \left(\sum_{t=k+1}^T \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t \right)^{-1}, \quad (3.5)$$

where a suitable estimator of $\boldsymbol{\Omega}$ will be discussed at the end of Section 3.2. Let us focus on the situation described in SCENARIO 2 when the lag of the model can change as well. Suppose that process $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ follows VAR(p) model up to time k and then changes to VAR(q), where $p < q$, with possibly different autoregressive

parameters. Situation when $p > q$ can be treated in a similar manner. Assuming that $p < q$, and using the analogous setup as above, the likelihood ratio test statistic is of the form $\max_{q < k \leq T} \Lambda_T^\dagger(k)$, where

$$\begin{aligned} \Lambda_T^\dagger(k) &:= \min_{\beta} \left\{ \sum_{t=p+1}^T (\mathbf{y}_t^* - \mathbf{M}_t^* \beta)^\top (\mathbf{y}_t^* - \mathbf{M}_t^* \beta) \right\} \\ &- \min_{\beta} \left\{ \sum_{t=p+1}^k (\mathbf{y}_t^* - \mathbf{M}_t^* \beta)^\top (\mathbf{y}_t^* - \mathbf{M}_t^* \beta) \right\} \\ &- \min_{\tilde{\beta}} \left\{ \sum_{t=k+1}^T (\mathbf{y}_t^* - \tilde{\mathbf{M}}_t^* \tilde{\beta})^\top (\mathbf{y}_t^* - \tilde{\mathbf{M}}_t^* \tilde{\beta}) \right\}, \end{aligned}$$

where the SCENARIO 2 and notation is described in (2.3).

In what follows we discuss the asymptotic behaviour of $\Lambda_T(k)$ for particular value of k and also $\Lambda_T := \max_{p < k < T} \{\Lambda_T(k)\}$ for SCENARIO 1 and $\Lambda_T^\dagger := \max_{p < k < T} \{\Lambda_T^\dagger(k)\}$ in case of SCENARIO 2. For the ease of notation it is more convenient to rescale the time interval on $(0, 1)$ and hence we define $Q_T(\tau) := \Lambda_T(\lfloor T\tau \rfloor)$ and $Q_T^\dagger(\tau) := \Lambda_T^\dagger(\lfloor T\tau \rfloor)$, $\tau \in (0, 1)$.

3.2 Tests based on the approximation by functionals of Wiener process

The main theorem is as follows:

Theorem 3.1 *Suppose that an n -dimensional stochastic process $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ follows a VAR(p) model of the form (2.2) and satisfies Assumptions A. Then, for test statistic (3.3) and under H_0*

$$Q_T(\tau) \xrightarrow[T \rightarrow \infty]{d} \frac{\|\mathbf{W}(\tau) - \tau \mathbf{W}(1)\|^2}{\tau(1 - \tau)}, \quad \tau \in (0, 1), \quad (3.6)$$

where \mathbf{W} is an $n(np + 1)$ -dimensional standard Wiener process with independent components.

Proof of the preceding theorem as well as other in this chapter is postponed to Section 3.4. Theorem 3.1 holds true also for Assumptions B and Assumptions C as will be proven later in more general case for unknown change-point τ . Term $\mathbf{W}(\tau) - \tau \mathbf{W}(1) =: \mathbf{B}(\tau)$ is $n(np + 1)$ -dimensional standard Brownian bridge process.

It follows from Theorem 3.1 that hypothesis H_0 of no change against the alternative under SCENARIO 1 that a change occurs at a given time k_0 is rejected on the level α if $Q_T(\tau_0) > C_\alpha$, where $\tau_0 = \frac{k_0}{T}$ and C_α is a critical value such that

$$\mathbb{P} \left[\frac{\|\mathbf{W}(\tau_0) - \tau_0 \mathbf{W}(1)\|^2}{\tau_0(1 - \tau_0)} \geq C_\alpha \right] = \alpha.$$

Next Theorem enables us to test H_0 against the alternative H_1 in SCENARIO 1 in case of unknown change point. This situation arises most often because data analysts usually do not have any information about the location of the break point.

Theorem 3.2 *Under Assumptions A and under H_0 , it holds that*

$$\sup_{\tau_1 < \tau < \tau_2} \{Q_T(\tau)\} \xrightarrow[T \rightarrow \infty]{d} \sup_{\tau_1 < \tau < \tau_2} \left\{ \frac{\|\mathbf{W}(\tau) - \tau \mathbf{W}(1)\|^2}{\tau(1-\tau)} \right\}, \quad (3.7)$$

for fixed $\tau_1, \tau_2 \in (0, 1)$.

There exist some approximations of the limiting distribution on the right-hand-side of (3.7) which can be used for computing the critical values. We will compare the approximation from James et al. (1987) with the simulation study. The limit distribution in (3.7) is influenced not only by the dimension of the Wiener process, but also by τ_1 and τ_2 . The choice of τ_1 and τ_2 is therefore not only a technical requirement. In particular, Andrews (1993) shows that if no restrictions are imposed (i.e. $\tau_1 = 0$ and $\tau_2 = 1$), the test diverges to infinity under H_0 . Simulations also show that the more observations we trim, the better power the testing procedure achieves. Weight functions can be utilized in case of early or late changes, see for instance Horváth (1993a), or Hušková et al. (2007) for the details and assumptions imposed on the weight functions.

The assumptions B or C can also be used instead of A. Any of these assumptions imply FCLTs.

Proposition 3.3 *Under Assumptions B or C and under H_0 , it holds that*

$$\sup_{\tau_1 < \tau < \tau_2} \{Q_T(\tau)\} \xrightarrow[T \rightarrow \infty]{d} \sup_{\tau_1 < \tau < \tau_2} \left\{ \frac{\|\mathbf{W}(\tau) - \tau \mathbf{W}(1)\|^2}{\tau(1-\tau)} \right\},$$

for fixed $\tau_1, \tau_2 \in (0, 1)$.

Let us move into the case of SCENARIO 2. The following result is an extension of Theorem 3.2:

Theorem 3.4 *Under H_0 , Assumptions A or B or C, if $p < q$, then for fixed $0 < \tau_1 < \tau_2 < 1$, it holds that*

$$\sup_{\tau_1 < \tau < \tau_2} \left\{ Q_T^\dagger(\tau) \right\} \xrightarrow[T \rightarrow \infty]{d} \max_{\tau_1 < \tau < \tau_2} \left\{ \frac{\|\mathbf{W}_1(\tau)\|^2}{\tau} + \frac{\|\mathbf{W}_1(1) - \mathbf{W}_1(\tau)\|^2}{1-\tau} - \|\mathbf{W}_1(1)\|^2 + \frac{\|\mathbf{W}_2(1) - \mathbf{W}_2(\tau)\|^2}{1-\tau} \right\}, \quad (3.8)$$

where $\mathbf{W}_1(\tau), \mathbf{W}_2(\tau)$ are two independent standard Wiener processes of dimensions $n(np+1)$ and $n^2(q-p)$, respectively.

As we have already pointed out, in practical applications the variance matrix $\mathbf{\Omega}$ is unknown. In that case we can replace it by a consistent estimate $\widehat{\mathbf{\Omega}}$ that satisfies $\widehat{\mathbf{\Omega}} - \mathbf{\Omega} = o_p(1)$. For example, if $\widehat{\boldsymbol{\beta}}_T$ is the least squares estimator of $\boldsymbol{\beta}$ in VAR(p) model (2.2) under H_0 , then the estimator of the form $\widehat{\mathbf{\Omega}}_T = \frac{1}{T-p} \sum_{t=p+1}^T \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t^\top$, where $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \mathbf{M}_t \widehat{\boldsymbol{\beta}}_T$ can be used due to Lemma 3.8. In addition, under Assumptions A and B there exists an almost sure rate of convergence for $\widehat{\mathbf{\Omega}}_T - \mathbf{\Omega}$, see Lemma 3.8.

3.3 Darling-Erdős-type test

In recent articles dealing with change points in multivariate models the authors considered only the approximation by certain functionals of Wiener processes like in (3.7). In this section, we will also discuss the convergence of the test statistics to the Gumbel-distributed random variable which has not been tackled by the previous authors in the context of multivariate VAR(p) models. Gumbel distribution is a double-exponential continuous distribution with distribution function $F(x; \mu, \beta) = \exp\{-e^{-(x-\mu)/\beta}\}$, where $\beta > 0$ and $\mu \in \mathbb{R}$ are the parameters. This distribution is widely used in the extreme value theory. As regards the change point problem in autoregressions, an approximation of the standardized quasi-likelihood ratio test statistic by the Gumbel-distributed random variable was derived in Davis et al. (1995) for *univariate* stationary autoregressive models. This section makes the extension to the stationary VAR(p) processes.

The advantage of this approach stands in the explicit calculation of asymptotic critical values hence they do not have to be simulated. The disadvantage lies in a relatively slow convergence, as it has often been pointed out in the literature, see for example Horváth (1993b), Davison (2003), or Hušková et al. (2007) among others.

Only SCENARIO 1 and Assumptions B will be discussed in this section. The extensions to other cases will be briefly discussed in the summary of the chapter.

Theorem 3.5 *Let us assume that the VAR(p) model satisfies Assumptions B. Then, under H_0 , it holds that*

$$\mathbb{P}\left[\frac{\Lambda_T - b_T(n(np+1))}{a_T(n(np+1))} \leq x\right] \xrightarrow{T \rightarrow \infty} \exp\left\{-2e^{-\frac{x}{2}}\right\}, \quad (3.9)$$

where

$$b_T(d) = \frac{\left(2 \ln \ln T + \frac{d}{2} \ln \ln \ln T - \ln \Gamma\left(\frac{d}{2}\right)\right)^2}{2 \ln \ln T},$$

$$a_T(d) = \sqrt{\frac{b_T(d)}{2 \ln \ln T}}$$

and $\Gamma(\cdot)$ is the gamma function.

The large values of the test statistic Λ_T show that the null hypothesis is violated.

The test statistic Λ_T depends on $\mathbf{\Omega}$. If matrix $\mathbf{\Omega}$ is not known, it can be replaced by its estimate $\hat{\mathbf{\Omega}}_T$ such that $\hat{\mathbf{\Omega}}_T - \mathbf{\Omega} = o_{\mathbb{P}}((\ln \ln T)^{-1})$, as $T \rightarrow \infty$. This condition is fulfilled for the least squares estimator due to Lemma 3.8 below. Note that Landau symbols used for vectors and matrices mean the boundedness (or convergence) rate of each element.

3.4 Proofs

One of the main tools for proving the statements in Chapter 3 is the invariance principle for martingale differences, strong mixing and m -dependent sequences,

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

see Theorems B.3, B.10 and B.11 in the Appendix B. The first lemma is a multivariate extension of Lemma 4.2. in Hušková et al. (2007) and it is necessary for proving uniform convergence results contained in Theorems 3.1 and 3.2 for the case of Assumptions A.

Lemma 3.6 *Let Assumptions A be fulfilled. Then, for any $0 < \gamma < \frac{1}{4}$, for all $\delta > 0$, there is $a = a(\delta)$ such that*

$$\mathbb{P} \left[\max_{p < k \leq T} \left\{ k^{\gamma-1} \cdot \left\| \mathbf{P}_k^{-1} - \mathbb{E}[\mathbf{P}_k^{-1}] \right\| \right\} > a \right] \leq \delta \quad (3.10)$$

and also

$$\mathbb{P} \left[\max_{p \leq k < T} \left\{ (T-k)^{\gamma-1} \cdot \left\| \tilde{\mathbf{P}}_k^{-1} - \mathbb{E}[\tilde{\mathbf{P}}_k^{-1}] \right\| \right\} > a \right] \leq \delta, \quad (3.11)$$

as $T \rightarrow \infty$, where \mathbf{P}_k and $\tilde{\mathbf{P}}_k$ are defined in (3.5).

Proof: Without loss of generality, let us assume that $\boldsymbol{\mu} = \mathbf{0}$ and let us begin with the proof of (3.10). Denote σ_{ij} an (i, j) -element of the matrix $\boldsymbol{\Omega}$ and σ^{ij} an (i, j) -element of the matrix $\boldsymbol{\Omega}^{-1}$. Matrix \mathbf{P}_k^{-1} (as well as $\tilde{\mathbf{P}}_k^{-1}$) consists of the sums of matrices $\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t$ which have $(np+1) \times (np+1)$ blocks and are block-symmetric. Denote the row number of the block by b_1 and the column number by b_2 . The first left-upper block (i.e. $b_1 = 1, b_2 = 1$) is of the form

$$\mathbf{B}_{1,1} = \begin{pmatrix} \sigma^{11} & \dots & \sigma^{1n} \\ \vdots & \ddots & \vdots \\ \sigma^{n1} & \dots & \sigma^{nn} \end{pmatrix} = \boldsymbol{\Omega}^{-1}.$$

The blocks \mathbf{B}_{b_1, b_2} for $b_1 = 1, b_2 > 1$ or $b_1 > 1, b_2 = 1$ respectively, are of the type $y_{i,t-s} \boldsymbol{\Omega}^{-1}$, $i = 1, \dots, n, s = 1, \dots, p$. The remaining blocks for $b_1 > 1, b_2 > 1$ are of the form $\mathbf{B}_{b_1, b_2} = y_{i,t-s} y_{j,t-u} \boldsymbol{\Omega}^{-1}$, $i, j = 1, \dots, n, s, u = 1, \dots, p$. Since $\mathbf{B}_{1,1} - \mathbb{E}[\mathbf{B}_{1,1}] = \mathbf{0}$, the conclusion for $\mathbf{B}_{1,1}$ is immediate.

For the other blocks we will apply an improvement of Hájek - Rényi inequality as given in Kokoszka and Leipus (2000): For all random variables η_1, \dots, η_T with finite second moments, for all non-negative c_1, \dots, c_T and for all $a > 0$

$$\begin{aligned} a^2 \mathbb{P} \left[\max_{1 \leq k \leq T} \left\{ c_k \left| \sum_{s=1}^k \eta_s \right| \right\} > a \right] &\leq \sum_{k=1}^{T-1} \left(|c_{k+1}^2 - c_k^2| \sum_{s=1}^k \sum_{t=1}^k \mathbb{E}[\eta_s \eta_t] \right) + \\ + 2 \cdot \sum_{k=1}^{T-1} \left(c_{k+1}^2 \mathbb{E} \left[|\eta_{k+1}| \left| \sum_{s=1}^k \eta_s \right| \right] \right) &+ \sum_{k=0}^{T-1} \left(c_{k+1}^2 \mathbb{E}[\eta_{k+1}^2] \right). \end{aligned} \quad (3.12)$$

Throughout this proof, $0 < K < \infty$ will be a generic constant (i.e. constant which can change from line to line). Let us begin the proof for blocks \mathbf{B}_{1, b_2} , $b_2 > 1$ and $\mathbf{B}_{b_1, 1}$, $b_1 > 1$. We apply inequality (3.12) with $\eta_s := \sigma^{\alpha\beta} y_{i,s}$, $i, \alpha, \beta = 1, \dots, n$ and $c_k := k^{\gamma-1}$.

The direct computation yields

$$\left| c_{k+1}^2 - c_k^2 \right| = \left| (k+1)^{2\gamma-2} - k^{2\gamma-2} \right| \leq 2(1-\gamma)k^{2\gamma-3}, \quad (3.13)$$

$$\begin{aligned} \mathbf{E} \left[y_{i_1, s-t_1} y_{i_2, s-t_2} \right] &= \sum_{\alpha_1=1}^n \sum_{\alpha_2=1}^n \sum_{u=0}^{\infty} \psi_{i_1, \alpha_1}^{(u)} \psi_{i_2, \alpha_2}^{(u+|t_1-t_2|)} \mathbf{E} \left[\varepsilon_{\alpha_1, s-t_1-u} \varepsilon_{\alpha_2, s-t_1-u} \right] = \\ &= \sum_{\alpha_1=1}^n \sum_{\alpha_2=1}^n \sigma_{\alpha_1, \alpha_2} \sum_{u=0}^{\infty} \psi_{i_1, \alpha_1}^{(u)} \psi_{i_2, \alpha_2}^{(u+|t_1-t_2|)}. \end{aligned}$$

According to Theorem 2.1 and (2.8), $\forall i_1, i_2 = 1, \dots, n, \forall t_1, t_2 = 1, \dots, p, \forall s = p+1, \dots, T$,

$$\left| \mathbf{E} \left[y_{i_1, s-t_1} y_{i_2, s-t_2} \right] \right| \leq K |t_1 - t_2|^{r-1} \cdot \lambda^{|t_1-t_2|},$$

where $0 < \lambda < 1$ is the largest solution in modulus of characteristic polynomial of $\text{VAR}(p)$ and r is the biggest multiplicity of these solutions. Since for any $r \in \mathbb{N}$ and $0 < \lambda < 1$

$$\sum_{s=p+1}^k \sum_{t=p+1}^k \left(|s-t|^{r-1} \cdot \lambda^{|s-t|} \right) = 2 \cdot \sum_{j=1}^{k-p-1} \left((k-p-j) \cdot j^{r-1} \cdot \lambda^j \right),$$

then

$$\begin{aligned} \left| \sum_{s=p+1}^k \sum_{t=p+1}^k \mathbf{E}[\eta_s \eta_t] \right| &\leq K \sum_{s=p+1}^k \sum_{t=p+1}^k \left| \mathbf{E} \left[y_{i_1, s} y_{i_2, t} \right] \right| = \\ &= K \sum_{j=1}^{k-p-1} \left(2(k-p-j) \cdot j^{r-1} \lambda^j \right) \leq K k, \quad (3.14) \end{aligned}$$

because series of the form $\sum_{j=1}^{\infty} j^{\zeta} \lambda^j$ is convergent for all $\zeta > 0$ and all $0 < \lambda < 1$. Since $\mathbf{E}[\eta_{k+1}^2] < K < \infty$, for some $K > 0$, then

$$\mathbf{E} \left[\left| \eta_{k+1} \right| \left| \sum_{s=p+1}^k \eta_s \right| \right] \leq \left(\mathbf{E} \eta_{k+1}^2 \right)^{\frac{1}{2}} \cdot \left(\sum_{s=p+1}^k \sum_{t=p+1}^k \mathbf{E}[\eta_s \eta_t] \right)^{\frac{1}{2}} \leq K \sqrt{k}. \quad (3.15)$$

Combining inequalities (3.13)–(3.15) with (3.12), one gets

$$\begin{aligned} &a^2 \mathbf{P} \left[\max_{1 \leq k \leq T} \left\{ k^{\gamma-1} \left| \sum_{s=1}^k \left(y_{i_1, s-t_1} y_{i_2, s-t_2} - \mathbf{E} \left[y_{i_1, s-t_1} y_{i_2, s-t_2} \right] \right) \right| \right\} > a \right] \leq \\ &\leq \sum_{k=1}^{T-1} \left(2(1-\gamma) k^{2\gamma-3} K k \right) + 2 \cdot \sum_{k=1}^{T-1} \left((k+1)^{2\gamma-2} K \sqrt{k} \right) + \sum_{k=1}^T k^{2\gamma-2} K \leq \\ &\leq K \left(\sum_{k=1}^T \frac{1}{k^{2-2\gamma}} + \sum_{k=2}^T \frac{1}{k^{\frac{3}{2}-2\gamma}} + \sum_{k=1}^T \frac{1}{k^{2-2\gamma}} \right) \leq \\ &\leq K \left(T^{2\gamma-1} + T^{2\gamma-\frac{1}{2}} + T^{2\gamma-1} \right) \xrightarrow{T \rightarrow \infty} 0, \end{aligned}$$

for $0 < \gamma < \frac{1}{4}$.

The proof of the desired inequality (3.10) for \mathbf{B}_{b_1, b_2} , $b_1 > 1$, $b_2 > 1$ is similar: Choose i_1, i_2, α, β arbitrarily between 1 and n , and t_1, t_2 between 1 and p and define $\eta_s := \sigma^{\alpha\beta} (y_{i_1, s-t_1} y_{i_2, s-t_2} - \mathbf{E} [y_{i_1, s-t_1} y_{i_2, s-t_2}])$ and $c_k := k^{\gamma-1}$.

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

Analogously, the direct computation yields

$$\left| \sum_{s=p+1}^k \sum_{t=p+1}^k \mathbf{E}[\eta_s \eta_t] \right| \leq K \sum_{s=p+1}^k \sum_{t=p+1}^k \left| \mathbf{E}[y_{i_1, s-t_1} y_{i_2, s-t_2} y_{i_3, t-t_1} y_{i_4, t-t_2}] - \mathbf{E}[y_{i_1, s-t_1} y_{i_2, s-t_2}] \mathbf{E}[y_{i_3, t-t_1} y_{i_4, t-t_2}] \right| \leq K k, \quad (3.16)$$

$$\mathbf{E} \left[|\eta_{k+1}| \left| \sum_{s=p+1}^k \eta_s \right| \right] \leq \left(\mathbf{E} \eta_{k+1}^2 \right)^{\frac{1}{2}} \cdot \left(\sum_{s=p+1}^k \sum_{t=p+1}^k \mathbf{E}[\eta_s \eta_t] \right)^{\frac{1}{2}} \leq K \sqrt{k}. \quad (3.17)$$

Combining inequalities (3.13), (3.16) and (3.17) with (3.12), one gets

$$a^2 \mathbf{P} \left[\max_{1 \leq k \leq T} \left\{ k^{\gamma-1} \left| \sum_{s=1}^k (y_{i_1, s-t_1} y_{i_2, s-t_2} - \mathbf{E}[y_{i_1, s-t_1} y_{i_2, s-t_2}]) \right| \right\} > a \right] \xrightarrow{T \rightarrow \infty} 0,$$

for $0 < \gamma < \frac{1}{4}$. The proof of (3.11) can be obtained in a similar way, if we use the fact that

$$\mathbf{P} \left[\max_{p \leq k < T} \left\{ (T-k)^{\gamma-1} \cdot \left| \sum_{t=k+1}^T \eta_t \right| \right\} > a \right] = \mathbf{P} \left[\max_{1 \leq k < T-p} \left\{ k^{\gamma-1} \cdot \left| \sum_{t=1}^k \eta_{T-t+1} \right| \right\} > a \right].$$

□

The next lemma is useful in the context of Assumptions C. It shows that under this set of conditions, products of the form $\varepsilon_{i,t} y_{j,t-s}$, $s > 0$, $i, j = 1, \dots, n$ fulfil similar conditions as those in (C.2) and (C.3) and hence FCLT B.11 holds. It is not important that the particular matrix $\mathbf{\Omega}$ is used, the statement of the lemma is valid for any non-random positive definite matrix.

Lemma 3.7 *Under Assumptions C and under H_0 , there exists $\mathbf{h} : \mathbb{R}^{n' \times \infty} \rightarrow \mathbb{R}^n$, a measurable function, such that for vector $\mathbf{m}_t := \mathbf{M}_t^\top \mathbf{\Omega}^{-1} \boldsymbol{\varepsilon}_t$, it holds*

$$\mathbf{m}_t = \mathbf{h}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots), \quad t \in \mathbb{Z}, \quad (3.18)$$

where $\{\boldsymbol{\nu}_t\}_{t \in \mathbb{Z}}$ is a sequence of iid random vectors with values in $\mathbb{R}^{n'}$. Further, there exists $m > 0$ and a sequence of $(2m+p)$ -dependent random vectors $\{\mathbf{m}_t^{(2m+p)}\}_{t \in \mathbb{Z}}$ such that $\mathbf{m}_t^{(2m+p)} = \mathbf{h}^{(2m+p)}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots, \boldsymbol{\nu}_{t-2m-p})$, $t \in \mathbb{Z}$, with measurable functions $\mathbf{h}^{(2m+p)} : \mathbb{R}^{n' \times (2m+p+1)} \rightarrow \mathbb{R}^n$, and

$$\sum_{m=1}^{\infty} \left(\mathbf{E} \left\| \mathbf{m}_0 - \mathbf{m}_0^{(2m+p)} \right\|^2 \right)^{\frac{1}{2}} < \infty. \quad (3.19)$$

Proof: According to Assumptions C and Theorem 2.2

$$\begin{aligned} \mathbf{m}_t &= (\mathbf{V}_t \otimes \mathbf{I}_n) \mathbf{\Omega}^{-1} \boldsymbol{\varepsilon}_t = \left((1, \mathbf{y}_{t-1}^\top, \dots, \mathbf{y}_{t-p}^\top)^\top \otimes \mathbf{I}_n \right) \mathbf{\Omega}^{-1} \boldsymbol{\varepsilon}_t \\ &= \left((1, \mathbf{g}^\top(\boldsymbol{\nu}_{t-1}, \dots), \dots, \mathbf{g}^\top(\boldsymbol{\nu}_{t-p}, \dots))^\top \otimes \mathbf{I}_n \right) \mathbf{\Omega}^{-1} \mathbf{f}(\boldsymbol{\nu}_t, \dots) =: \\ &=: \mathbf{h}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots), \quad t \in \mathbb{Z}, \end{aligned}$$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

where \mathbf{h} is measurable and hence (3.18) holds. Let us denote

$$\mathbf{V}_t^{(2m)\top} := (1, \mathbf{y}_{t-1}^{(2m)\top}, \dots, \mathbf{y}_{t-p}^{(2m)\top}), \quad t \in \mathbb{Z},$$

where $\mathbf{y}_t^{(2m)}$'s are defined in (2.18), and let

$$\mathbf{m}_t^{(2m+p)} := \left(\mathbf{V}_t^{(2m)} \otimes \mathbf{I}_n \right) \Omega^{-1} \boldsymbol{\varepsilon}_t^{(m)}.$$

Then due to (2.18)

$$\begin{aligned} \mathbf{m}_t^{(2m+p)} &= \left(\left(1, \mathbf{g}^{(2m)\top}(\boldsymbol{\nu}_{t-1}, \dots, \boldsymbol{\nu}_{t-1-2m}), \dots, \right. \right. \\ &\quad \left. \left. \dots, \mathbf{g}^{(2m)\top}(\boldsymbol{\nu}_{t-p}, \dots, \boldsymbol{\nu}_{t-p-2m}) \right)^\top \otimes \mathbf{I}_n \right) \cdot \Omega^{-1} \mathbf{f}^{(m)}(\boldsymbol{\nu}_t, \dots, \boldsymbol{\nu}_{t-m}) =: \\ &=: \mathbf{h}^{(2m+p)}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots, \boldsymbol{\nu}_{t-2m-p}), \end{aligned}$$

where $\mathbf{h}^{(2m+p)}$ are measurable functions and hence we established $(2m+p)$ -dependence.

Let us check the validity of (3.19):

$$\begin{aligned} &\sum_{m=1}^{\infty} \left(\mathbb{E} \left[\mathbf{m}_0 - \mathbf{m}_0^{(2m+p)} \right]^2 \right)^{\frac{1}{2}} = \sum_{m=1}^{\infty} \nu_2 \left(\mathbf{m}_0 - \mathbf{m}_0^{(2m+p)} \right) = \\ &\leq \sum_{m=1}^{\infty} \nu_2 \left((\mathbf{V}_0 \otimes \mathbf{I}_n) \Omega^{-1} (\boldsymbol{\varepsilon}_0 - \boldsymbol{\varepsilon}_0^{(m)}) \right) + \\ &\quad + \sum_{m=1}^{\infty} \nu_2 \left(\left((\mathbf{V}_0 - \mathbf{V}_0^{(2m)}) \otimes \mathbf{I}_n \right) \Omega^{-1} (\boldsymbol{\varepsilon}_0^{(m)} - \boldsymbol{\varepsilon}_0) \right) + \\ &\quad + \sum_{m=1}^{\infty} \nu_2 \left(\left((\mathbf{V}_0 - \mathbf{V}_0^{(2m)}) \otimes \mathbf{I}_n \right) \Omega^{-1} \boldsymbol{\varepsilon}_0 \right) \leq \\ &\leq K_1 \sum_{m=1}^{\infty} \nu_2 \left(\boldsymbol{\varepsilon}_0 - \boldsymbol{\varepsilon}_0^{(m)} \right) + K_2 \left(\sum_{m=1}^{\infty} \nu_2^2 \left(\boldsymbol{\varepsilon}_0 - \boldsymbol{\varepsilon}_0^{(m)} \right) \right)^{\frac{1}{2}} + \\ &\quad + K_3 \left(\sum_{m=1}^{\infty} \nu_2^2 \left(\mathbf{y}_0 - \mathbf{y}_0^{(2m)} \right) \right)^{\frac{1}{2}} + K_4 \sum_{m=1}^{\infty} \nu_2 \left(\mathbf{y}_0 - \mathbf{y}_0^{(2m)} \right) < \infty, \end{aligned}$$

due to (2.11) and (2.13), with K_1, K_2, K_3, K_4 being finite positive constants. \square

The upcoming Lemma 3.8 treats the speed of convergence of the least squares estimators to the true values. Stronger consistency results can be stated in case of Assumptions A or B since in these situations there is a guaranteed speed in FCLT. Hence, in the context of this lemma, Assumptions C can be viewed as somewhat weaker compared to A or B. As regards the proof of consistency under Assumptions B or C, it might seem non-standard for the reader to prove it using FCLT, however in this way, we can build up on the previously proven statements in Chapter 2 and on Lemma 3.7. The proof can hence be shorter.

Lemma 3.8 *Under Assumptions A or B it holds that*

$$\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} = \mathcal{O} \left(\sqrt{\frac{\ln \ln T}{T}} \right) \quad \text{a.s.}, \quad (3.20)$$

$$\widehat{\boldsymbol{\Omega}}_T - \boldsymbol{\Omega} = \mathcal{O} \left(\sqrt{\frac{\ln \ln T}{T}} \right) \quad \text{a.s.} \quad (3.21)$$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

as $T \rightarrow \infty$.

Under Assumptions C, $\widehat{\beta}_T - \beta = o_P(1)$, $\widehat{\Omega}_T - \Omega = o_P(1)$, as $T \rightarrow \infty$.

Proof: (a) Proof under Assumptions A:

In case of (3.20), it is sufficient to check Assumptions 2.1 and 2.2 presented in Nielsen (2005). First, $\sup_t \mathbf{E}[\|\boldsymbol{\varepsilon}_t\|^{2+\delta} | \mathfrak{F}_{t-1}] < \infty$ a.s., $\delta > 0$, is fulfilled due to condition (A.2). Assumption $\liminf_{t \rightarrow \infty} \lambda_{\min}(\mathbf{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top | \mathfrak{F}_{t-1}]) > 0$ a.s. on the smallest eigenvalue is fulfilled as well due to the assumption of positive-definiteness of the conditional variance matrix Ω . Hence according to Nielsen (2005), Theorem 2.5, (3.20) is fulfilled.

Since Assumption 2.7 in Nielsen (2005) of constant positive definite conditional variance of $\boldsymbol{\varepsilon}_t$ is fulfilled too, then according to Nielsen (2005), Corollary 2.9

$$\widehat{\Omega}_T - \Omega = o(T^{-\eta}) \quad \text{a.s.}, \quad T \rightarrow \infty, \quad \forall \eta < \frac{1}{2},$$

hence (3.21) holds true as well.

(b) Proof under Assumptions B:

The left-hand-side of (3.20) is

$$\widehat{\beta}_T - \beta = \left(\frac{1}{T} \sum_{t=p+1}^T \mathbf{M}_t^\top \mathbf{M}_t \right)^{-1} \cdot \left(\frac{1}{T} \sum_{t=p+1}^T \mathbf{M}_t^\top \boldsymbol{\varepsilon}_t \right). \quad (3.22)$$

It holds that $\mathbf{E}[\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t] = \mathbf{0}$, and for $s > 0$

$$\text{vec } \mathbf{E}[\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-s}^\top \mathbf{M}_{t-s}] = \mathbf{E}[\mathbf{M}_{t-s}^\top \boldsymbol{\varepsilon}_{t-s} \otimes \mathbf{M}_t^\top] \text{vec } \mathbf{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0},$$

where we used rule (1) in Lemma A.1. Using the latter rule again, we have

$$\begin{aligned} \text{vec var}(\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t) &= \mathbf{E}[\text{vec}(\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \mathbf{M}_t)] = \mathbf{E}[\mathbf{M}_t^\top \otimes \mathbf{M}_t^\top] \cdot \text{vec}(\mathbf{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top]) = \\ &= \mathbf{E}[\mathbf{M}_t^\top \otimes \mathbf{M}_t^\top] \cdot \text{vec}(\Omega) = \text{vec}(\mathbf{E}[\mathbf{M}_t^\top \Omega \mathbf{M}_t]) =: \text{vec}(\Delta). \end{aligned}$$

Hence process $\{\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ is centered and weakly stationary with finite variance. In addition, $\sup_t \mathbf{E} \|\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t\|^{2+\delta} = K < \infty$, $\delta > 0$, for some positive K . Since $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \mathbf{c} - \Phi_1 \mathbf{y}_{t-1} - \dots - \Phi_p \mathbf{y}_{t-p}$, then $\boldsymbol{\varepsilon}_t$ is a measurable function of finitely many strong mixing vectors $\mathbf{y}_t, \dots, \mathbf{y}_{t-p}$. According to Theorem B.9 we obtain that $\{\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ is a strong mixing process with the same size as $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$. According to Theorem B.10, there exists $n(np+1)$ -dimensional Wiener process \mathbf{W}_Δ with variance matrix Δ , such that

$$\sum_{t=p+1}^T \mathbf{M}_t^\top \boldsymbol{\varepsilon}_t - \mathbf{W}_\Delta(T) = \mathcal{O}(T^{\frac{1}{2}-\lambda}) \quad \text{a.s.}, \quad T \rightarrow \infty,$$

for some $\lambda > 0$. Dividing by T we get

$$\frac{1}{T} \sum_{t=p+1}^T \mathbf{M}_t^\top \boldsymbol{\varepsilon}_t = \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \quad \text{a.s.}, \quad T \rightarrow \infty, \quad (3.23)$$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

where we used the law of the iterated logarithm for the Wiener process $\mathbf{W}_\Delta(T) = \mathcal{O}(\sqrt{T \ln \ln T})$ a.s., as $T \rightarrow \infty$.

Similarly, vector $\left\{ \text{vec}(\mathbf{M}_t^\top \mathbf{M}_t - \mathbb{E}[\mathbf{M}_t^\top \mathbf{M}_t]) \right\}_{t \in \mathbb{Z}}$ satisfies conditions in Theorem B.10, and hence

$$\frac{1}{T} \sum_{t=p+1}^T \mathbf{M}_t^\top \mathbf{M}_t = \mathbb{E}[\mathbf{M}_1^\top \mathbf{M}_1] + \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \quad \text{a.s., } T \rightarrow \infty. \quad (3.24)$$

Plugging in results (3.23) and (3.24) into (3.22), we obtain (3.20).

For the estimator of variance matrix we can write

$$\begin{aligned} \text{vec}(\widehat{\Omega}_T - \Omega) &= \frac{1}{T-p} \cdot \text{vec} \left(\sum_{t=p+1}^T [\mathbf{y}_t - \mathbf{M}_t \widehat{\beta}_T] [\mathbf{y}_t - \mathbf{M}_t \widehat{\beta}_T]^\top \right) - \\ &\quad - \text{vec}(\Omega) = \\ &= \frac{1}{T-p} \text{vec} \left(\sum_{t=p+1}^T [\boldsymbol{\varepsilon}_t - \mathbf{M}_t (\widehat{\beta}_T - \beta)] [\boldsymbol{\varepsilon}_t - \mathbf{M}_t (\widehat{\beta}_T - \beta)]^\top \right) - \text{vec}(\Omega) = \\ &= \frac{1}{T-p} \sum_{t=p+1}^T \text{vec} [\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \Omega] - \end{aligned} \quad (3.25)$$

$$- \frac{1}{T-p} \sum_{t=p+1}^T \text{vec} [\mathbf{M}_t (\widehat{\beta}_T - \beta) \boldsymbol{\varepsilon}_t^\top] + \quad (3.26)$$

$$+ \frac{1}{T-p} \sum_{t=p+1}^T \text{vec} [\boldsymbol{\varepsilon}_t (\widehat{\beta}_T - \beta)^\top \mathbf{M}_t^\top] - \quad (3.27)$$

$$- \frac{1}{T-p} \sum_{t=p+1}^T \text{vec} [\mathbf{M}_t (\widehat{\beta}_T - \beta) (\widehat{\beta}_T - \beta)^\top \mathbf{M}_t^\top]. \quad (3.28)$$

Let us start with (3.25) and define $\boldsymbol{\xi}_t := \text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \Omega) = \widetilde{\mathbf{f}}(\mathbf{y}_t, \dots, \mathbf{y}_{t-p})$ which is the measurable function of finitely many strong mixing vectors and due to Theorem B.9 is a strong mixing vector with the same rate as the strong mixing process $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$. It can be easily checked that process $\{\boldsymbol{\xi}_t\}_{t \in \mathbb{Z}}$ is centered and weakly stationary and satisfies other conditions of Theorem B.10 as well. Due to this theorem and the Law of the iterated logarithm for the Wiener process we obtain that term (3.25) is $\mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right)$ a.s., as $T \rightarrow \infty$.

Term (3.26) can be rewritten using rule (1) in Lemma A.1 as

$$\frac{1}{T-p} \sum_{t=p+1}^T \text{vec} [\mathbf{M}_t (\widehat{\beta}_T - \beta) \boldsymbol{\varepsilon}_t^\top] = \left(\frac{1}{T-p} \sum_{t=p+1}^T (\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t) \right) (\widehat{\beta}_T - \beta). \quad (3.29)$$

$\text{vec}(\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t)$ is centered weakly stationary strong mixing sequence satisfying conditions of Theorem B.10 and hence

$$\frac{1}{T-p} \sum_{t=p+1}^T (\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t) = \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \quad \text{a.s., } T \rightarrow \infty.$$

Due to (3.20), (3.26) is $\mathcal{O}\left(\frac{\ln \ln T}{T}\right)$ a.s., as $T \rightarrow \infty$. Term (3.27) can be treated in the same way.

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

Due to (3.20) and (3.24), we get for term (3.28)

$$\begin{aligned} & \frac{1}{T-p} \sum_{t=p+1}^T \text{vec} \left[\mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right] = \\ & = \frac{1}{T-p} \cdot \sum_{t=p+1}^T (\mathbf{M}_t \otimes \mathbf{M}_t) \text{vec} \left((\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \right) = \mathcal{O} \left(\frac{\ln \ln T}{T} \right) \quad \text{a.s.}, \end{aligned}$$

as $T \rightarrow \infty$. By combining the approximations for terms (3.25)–(3.28) together, we get the assertion.

(c) Proof under Assumptions C

Let us begin with sequence $\{\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$. Due to (C.1) and (C.4), we get that $\mathbb{E}[\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t] = \mathbf{0}$ and $\mathbb{E} \|\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t\|^2 < \infty$, for all $t \in \mathbb{Z}$. Due to Lemma 3.7 with $\boldsymbol{\Omega} = \mathbf{I}_n$, we can apply Theorem B.11 and hence

$$\frac{1}{\sqrt{T}} \cdot \sum_{t=1}^{\lfloor T\tau \rfloor} \mathbf{M}_t^\top \boldsymbol{\varepsilon}_t \xrightarrow[T \rightarrow \infty]{\text{d}} \mathbf{W}_\Delta(\tau), \quad \tau \in (0, 1),$$

where $\Delta = \text{var}[\mathbf{M}_t^\top \boldsymbol{\varepsilon}_t]$ due to strict stationarity of $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ and due to (C.4). From that we obtain

$$\frac{1}{T} \sum_{t=p+1}^T \mathbf{M}_t^\top \boldsymbol{\varepsilon}_t = o_P(1), \quad T \rightarrow \infty. \quad (3.30)$$

Theorem B.12 implies

$$\frac{1}{\sqrt{T}} \cdot \sum_{t=p+1}^{\lfloor T\tau \rfloor} \left(\text{vech}(\mathbf{M}_t^\top \mathbf{M}_t) - \mathbb{E}[\text{vech}(\mathbf{M}_t^\top \mathbf{M}_t)] \right) \xrightarrow[T \rightarrow \infty]{\text{d}} \mathbf{W}_\mathbf{T}(\tau),$$

where $\mathbf{T} := \text{var}[\text{vech}(\mathbf{M}_1^\top \mathbf{M}_1)]$. Since $\mathbb{E}[\text{vech}(\mathbf{M}_t^\top \mathbf{M}_t)]$ is independent of t we get that

$$\frac{1}{T} \sum_{t=p+1}^T \mathbf{M}_t^\top \mathbf{M}_t = \mathbb{E}[\mathbf{M}_1^\top \mathbf{M}_1] + o_P(1), \quad T \rightarrow \infty. \quad (3.31)$$

Substituting (3.30) and (3.31) into (3.22) we get $\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} = o_P(1)$, as $T \rightarrow \infty$.

As regards consistency of $\hat{\boldsymbol{\Omega}}_T$ we will use expressions (3.25)–(3.28). Consistency of (3.25) follows directly from Theorem B.12. As regards terms (3.26) and (3.27), we can use (3.29), (3.30) and consistency of $\hat{\boldsymbol{\beta}}_T$. As regards the consistency of (3.28) it suffices to use (3.31) and consistency of $\hat{\boldsymbol{\beta}}_T$. \square

Proof of Theorem 3.1: Let us start with $\Lambda_T(k)$ as given in (3.3). Since $\mathbf{s}_T - \mathbf{s}_k = \tilde{\mathbf{s}}_k$, we have

$$\Lambda_T(k) = -\mathbf{s}_T^\top \mathbf{P}_T \mathbf{s}_T + \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k + (\mathbf{s}_T - \mathbf{s}_k)^\top \tilde{\mathbf{P}}_k (\mathbf{s}_T - \mathbf{s}_k), \quad (3.32)$$

where the notation is defined in (3.4) and (3.5).

We will treat each summand in (3.32) separately. Let us focus on the vector \mathbf{s}_T appearing in the first addend. Vectors $\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t$ are elements of an $r = n(np+1)$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

dimensional sequence of martingale differences with respect to \mathfrak{F}_t . Let us denote $\Upsilon := \mathbb{E} \left[\mathbf{M}_t^\top \Omega^{-1} \mathbf{M}_t \right] \in \mathbb{R}^{n(np+1) \times n(np+1)}$ which is correctly defined because under H_0 , matrix $\mathbb{E} \left[\mathbf{M}_t^\top \Omega^{-1} \mathbf{M}_t \right]$ does not depend on t . Then,

$$\mathbf{u}_{T,t} := \frac{1}{\sqrt{T-p}} \Upsilon^{-\frac{1}{2}} \mathbf{M}_t^\top \Omega^{-1} \boldsymbol{\varepsilon}_t$$

is a martingale difference array. It holds $\mathbb{E}[\mathbf{u}_{T,t}] = \mathbf{0}$ and

$$\mathbb{E} \left[\mathbf{u}_{T,t} \mathbf{u}_{T,t}^\top \right] = \frac{1}{T-p} \Upsilon^{-\frac{1}{2}} \mathbb{E} \left[\mathbf{M}_t^\top \Omega^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \Omega^{-1} \mathbf{M}_t \right] \Upsilon^{-\frac{1}{2}} = \frac{1}{T-p} \mathbf{I}_r, \quad (3.33)$$

because the expectation inside (3.33) is

$$\mathbb{E} \left[\mathbb{E} \left[\mathbf{M}_t^\top \Omega^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \Omega^{-1} \mathbf{M}_t \mid \mathfrak{F}_{t-1} \right] \right] = \mathbb{E} \left[\mathbf{M}_t^\top \Omega^{-1} \mathbb{E} \left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \mid \mathfrak{F}_{t-1} \right] \Omega^{-1} \mathbf{M}_t \right] = \Upsilon.$$

Similarly as in Hušková et al. (2007), it will be shown that the conditions in Theorem B.3, are fulfilled:

$$\sum_{t=p+1}^T \mathbf{u}_{T,t} \mathbf{u}_{T,t}^\top \xrightarrow[T \rightarrow \infty]{\mathbb{P}} \mathbf{I}_r, \quad (3.34)$$

$$\max_{p+1 \leq t \leq T} \left\{ \mathbf{u}_{T,t}^\top \mathbf{u}_{T,t} \right\} \xrightarrow[T \rightarrow \infty]{\mathbb{P}} 0, \quad (3.35)$$

$$\lim_{T \rightarrow \infty} \sum_{t=p+1}^{\lfloor T\tau \rfloor} \text{var} \mathbf{u}_{T,t} = \tau \mathbf{I}_r, \quad \forall \tau \in [0, 1]. \quad (3.36)$$

First, (3.34) will be shown. Due to (3.33), statement (3.34) is equivalent to

$$\sum_{t=p+1}^T \mathbf{u}_{T,t} \mathbf{u}_{T,t}^\top - \mathbb{E} \left[\sum_{t=p+1}^T \mathbf{u}_{T,t} \mathbf{u}_{T,t}^\top \right] \xrightarrow[T \rightarrow \infty]{\mathbb{P}} \mathbf{0}. \quad (3.37)$$

The left-hand side of (3.37) is of the form

$$\Upsilon^{-\frac{1}{2}} \cdot \left\{ \frac{1}{T-p} \sum_{t=p+1}^T \left[\mathbf{M}_t^\top \Omega^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \Omega^{-1} \mathbf{M}_t - \mathbb{E} \left[\mathbf{M}_t^\top \Omega^{-1} \mathbf{M}_t \right] \right] \right\} \Upsilon^{-\frac{1}{2}}.$$

The formula inside the curly brackets can be divided into two terms:

$$\frac{1}{T-p} \sum_{t=p+1}^T \left(\mathbf{M}_t^\top \Omega^{-1} \left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \Omega \right] \Omega^{-1} \mathbf{M}_t \right) + \quad (3.38)$$

$$\frac{1}{T-p} \sum_{t=p+1}^T \left(\mathbf{M}_t^\top \Omega^{-1} \mathbf{M}_t - \mathbb{E} \left[\mathbf{M}_t^\top \Omega^{-1} \mathbf{M}_t \right] \right). \quad (3.39)$$

The summands in (3.38) are elements of a strictly stationary and ergodic sequence which can be written as

$$\text{vec} \left[\mathbf{M}_t^\top \Omega^{-1} (\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \Omega) \Omega^{-1} \mathbf{M}_t \right] = \left((\mathbf{M}_t^\top \Omega^{-1}) \otimes (\mathbf{M}_t^\top \Omega^{-1}) \right) \cdot \text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \Omega).$$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

The sequence has zero expectation since

$$\begin{aligned} & \mathbb{E} \left[\left((\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1}) \otimes (\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1}) \right) \cdot \text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \boldsymbol{\Omega}) \right] = \\ &= \mathbb{E} \left[\mathbb{E} \left[\left((\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1}) \otimes (\boldsymbol{\Omega}^{-1} \mathbf{M}_t) \right) \cdot \text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \boldsymbol{\Omega}) \middle| \mathfrak{F}_{t-1} \right] \right] = \\ &= \mathbb{E} \left[\left(\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \right) \otimes \left(\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \right) \cdot \text{vec} \left(\mathbb{E} \left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \boldsymbol{\Omega} \middle| \mathfrak{F}_{t-1} \right] \right) \right] = \mathbf{0}. \end{aligned}$$

Applying the Ergodic Theorem B.7, we get

$$\frac{1}{T-p} \cdot \sum_{t=p+1}^T \left(\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \boldsymbol{\Omega} \right] \boldsymbol{\Omega}^{-1} \mathbf{M}_t \right) \xrightarrow[T \rightarrow \infty]{\mathbb{P}} \mathbf{0}.$$

The expression (3.39) goes to zero in probability, due to Theorem B.7 as well.

Now, we will show that (3.35) is satisfied. By using Chebyshev's inequality we have

$$\begin{aligned} \mathbb{P} \left[\max_{p+1 \leq t \leq T} \{ \mathbf{u}_{T,t}^\top \mathbf{u}_{T,t} \} > \eta \right] &\leq \sum_{t=p+1}^T \mathbb{P} \left[\frac{1}{T-p} \cdot \boldsymbol{\varepsilon}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t \boldsymbol{\Upsilon}^{-1} \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t > \eta \right] \leq \\ &\leq \frac{1}{\eta^2 (T-p)^2} \cdot \sum_{t=p+1}^T \mathbb{E} \left[\boldsymbol{\varepsilon}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t \boldsymbol{\Upsilon}^{-1} \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \right]^2 \\ &\xrightarrow[T \rightarrow \infty]{} 0, \end{aligned}$$

since due to Lemma A.2

$$\begin{aligned} \mathbb{E} \left[\boldsymbol{\varepsilon}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t \boldsymbol{\Upsilon}^{-1} \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \right]^2 &= \text{tr} \left\{ \mathbb{E} \left[\text{vec}(\boldsymbol{\Omega}^{-1} \mathbf{M}_t \boldsymbol{\Upsilon}^{-1} \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1}) \cdot \right. \right. \\ &\quad \left. \left. \text{vec}(\boldsymbol{\Omega}^{-1} \mathbf{M}_t \boldsymbol{\Upsilon}^{-1} \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1})^\top \right] \cdot \mathbb{E} \left[(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) \middle| \mathfrak{F}_{t-1} \right] \right\} < K < \infty, \end{aligned}$$

where K is a positive constant independent of t .

The proof of (3.36) is simple, because from (3.33)

$$\lim_{T \rightarrow \infty} \sum_{t=p+1}^{\lfloor T\tau \rfloor} \text{var} \mathbf{u}_{T,t} = \lim_{T \rightarrow \infty} \sum_{t=p+1}^{\lfloor T\tau \rfloor} \frac{1}{T-p} \mathbf{I}_r = \lim_{T \rightarrow \infty} \frac{\lfloor T\tau \rfloor}{T} \mathbf{I}_r = \tau \mathbf{I}_r.$$

It follows from Theorem B.3 that

$$\frac{1}{\sqrt{T-p}} \cdot \boldsymbol{\Upsilon}^{-\frac{1}{2}} \cdot \sum_{t=p+1}^{\lfloor T\tau \rfloor} \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \xrightarrow[T \rightarrow \infty]{\text{d}} \mathbf{W}(\tau),$$

and, alternatively, $\forall \tau \in [0, 1]$

$$\frac{1}{\sqrt{T}} \cdot \boldsymbol{\Upsilon}^{-\frac{1}{2}} \cdot \mathbf{s}_{\lfloor T\tau \rfloor} \xrightarrow[T \rightarrow \infty]{\text{d}} \mathbf{W}(\tau). \quad (3.40)$$

Term $\tilde{\mathbf{s}}_k$ can be treated in the same way.

Let us focus on the matrix \mathbf{P}_T . From Lemma 3.6, we obtain

$$\frac{1}{T-p} \mathbf{P}_T^{-1} = \frac{1}{T-p} \sum_{t=p+1}^T \left(\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t \right) \xrightarrow[T \rightarrow \infty]{\mathbb{P}} \mathbb{E}[\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t] = \boldsymbol{\Upsilon}$$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

from which we immediately get that $\sqrt{T} \mathbf{P}_T^{\frac{1}{2}} \xrightarrow[T \rightarrow \infty]{\mathbf{P}} \mathbf{\Upsilon}^{-\frac{1}{2}}$. Combining this with (3.40) and using Hamilton (1994), p. 184, Proposition 7.3(b) we obtain $\mathbf{P}_T^{\frac{1}{2}} \cdot \mathbf{s}_T \xrightarrow[T \rightarrow \infty]{\mathbf{d}} \mathbf{W}(1)$. Hence

$$-\mathbf{s}_T^\top \mathbf{P}_T \mathbf{s}_T = -\left(\mathbf{P}_T^{\frac{1}{2}} \mathbf{s}_T\right)^\top \cdot \left(\mathbf{P}_T^{\frac{1}{2}} \mathbf{s}_T\right) \xrightarrow[T \rightarrow \infty]{\mathbf{d}} -\|\mathbf{W}(1)\|^2. \quad (3.41)$$

The asymptotic behaviour of the first addend in the test statistic $Q_T(\tau)$ has been developed. It remains to investigate asymptotic behaviour of matrices \mathbf{P}_k and $\tilde{\mathbf{P}}_k$. In view of Lemma 3.6 it holds

$$\begin{aligned} k^{-1} \left(\mathbf{P}_k^{-1} - \mathbb{E}[\mathbf{P}_k^{-1}] \right) &= \mathcal{O}_{\mathbf{P}}(T^{-\gamma}), \\ (T-k)^{-1} \left(\tilde{\mathbf{P}}_k^{-1} - \mathbb{E}[\tilde{\mathbf{P}}_k^{-1}] \right) &= \mathcal{O}_{\mathbf{P}}(T^{-\gamma}), \end{aligned}$$

uniformly for $p < k < T$ and for $0 < \gamma < \frac{1}{4}$. Hence, $T \mathbf{P}_{[T\tau]} \xrightarrow[T \rightarrow \infty]{\mathbf{P}} (\tau \mathbf{\Upsilon})^{-1}$. Combining the latest result with (3.40) we get

$$\mathbf{s}_{[T\tau]}^\top \mathbf{P}_{[T\tau]} \mathbf{s}_{[T\tau]} \xrightarrow[T \rightarrow \infty]{\mathbf{d}} \frac{\|\mathbf{W}(\tau)\|^2}{\tau}. \quad (3.42)$$

Analogously,

$$\left(\mathbf{s}_T - \mathbf{s}_{[T\tau]}\right)^\top \tilde{\mathbf{P}}_{[T\tau]} \left(\mathbf{s}_T - \mathbf{s}_{[T\tau]}\right) \xrightarrow[T \rightarrow \infty]{\mathbf{d}} \frac{\|\mathbf{W}(1) - \mathbf{W}(\tau)\|^2}{1 - \tau}. \quad (3.43)$$

Combining (3.41)–(3.43) and using continuous mapping theorem we obtain

$$Q_T(\tau) \xrightarrow[T \rightarrow \infty]{\mathbf{d}} \frac{\|\mathbf{W}(\tau) - \tau \mathbf{W}(1)\|^2}{\tau(1 - \tau)}$$

which concludes the proof. \square

Proof of Theorem 3.2: The proof follows from Theorem 3.1, continuity of the supremum in the Skorohod space $D(0, 1)$ and in the continuous mapping theorem, see for example Billingsley (1999), page 29. \square

Proof of Proposition 3.3:

Proof under Assumptions B:

The proof is analogous to the proof presented in Davis et al. (1995), the key steps can be seen in the proof of Lemma 3.9.

Proof under Assumptions C:

It follows from part (c) of the proof of Lemma 3.8 that for $\{\mathbf{m}_t := (\mathbf{M}_t^\top \mathbf{\Omega}^{-1} \boldsymbol{\varepsilon}_t)\}_{t \in \mathbb{Z}}$ and for $\{\mathbf{w}_t := \text{vech}(\mathbf{M}_t^\top \mathbf{\Omega}^{-1} \mathbf{M}_t)\}_{t \in \mathbb{Z}}$, we get

$$\begin{aligned} \frac{1}{\sqrt{T}} \cdot \sum_{t=p+1}^{[T\tau]} \mathbf{m}_t &\xrightarrow[T \rightarrow \infty]{\mathbf{d}} \mathbf{W}_{\mathbf{\Upsilon}}(\tau), \\ \frac{1}{\sqrt{T}} \cdot \sum_{t=p+1}^{[T\tau]} \left(\mathbf{w}_t - \mathbb{E}[\mathbf{w}_t]\right) &\xrightarrow[T \rightarrow \infty]{\mathbf{d}} \mathbf{W}_{\mathbf{\Sigma}}(\tau), \end{aligned} \quad (3.44)$$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

where $\Sigma = \text{var}[\mathbf{w}_t]$ for all t , $\mathbf{E}[\mathbf{w}_t] = \mathbf{E}[\text{vech}(\mathbf{M}_1^\top \Omega^{-1} \mathbf{M}_1)]$ for all t , and \mathbf{W}_Υ is $n(np+1)$ -dimensional Wiener process with variance matrix Υ and \mathbf{W}_Σ is $\binom{n(np+1)[n(np+1)+1]}{2}$ -dimensional Wiener process with variance matrix Σ . Using (3.44) it can be established that

$$\begin{aligned} ([T\tau] \mathbf{P}_{[T\tau]}) &\xrightarrow[T \rightarrow \infty]{\mathbf{P}} \Upsilon^{-1}, \\ ((T - [T\tau]) \tilde{\mathbf{P}}_{[T\tau]}) &\xrightarrow[T \rightarrow \infty]{\mathbf{P}} \Upsilon^{-1}, \end{aligned}$$

uniformly for $\tau \in (0, 1)$. From that we easily obtain the results (3.41)–(3.43) and hence the conclusion is immediate. \square

Proof of Theorem 3.4: As in case of SCENARIO 1, the key is to obtain asymptotic distribution of all three terms of the test statistic. For Assumptions A, this can be done by using similar arguments as in the proof of Theorem 3.2. In case of Assumptions B or C, the proof of Proposition 3.3 can be used analogically. Hence, under Assumptions A or B or C it can easily be established that

$$Q_T^\dagger([T\tau]) \xrightarrow[T \rightarrow \infty]{\mathbf{d}} L(\tau),$$

where

$$\begin{aligned} L(\tau) &= -\mathbf{W}_p^*(1)^\top \Upsilon_p^{-1} \mathbf{W}_p^*(1) + \frac{\mathbf{W}_p^*(\tau)^\top \Upsilon_p^{-1} \mathbf{W}_p^*(\tau)}{\tau} + \\ &+ \frac{(\mathbf{W}_q^*(1) - \mathbf{W}_q^*(\tau))^\top \Upsilon_q^{-1} (\mathbf{W}_q^*(1) - \mathbf{W}_q^*(\tau))}{1 - \tau} \end{aligned} \quad (3.45)$$

and $\mathbf{W}_q^*(\tau)$ is an $n(nq+1)$ -dimensional Wiener process with variance matrix $\Upsilon_q = \mathbf{E}[\tilde{\mathbf{M}}_t^\top \Omega^{-1} \tilde{\mathbf{M}}_t] \in \mathbb{R}^{n(nq+1) \times n(nq+1)}$ and $\mathbf{W}_p^*(\tau)$ is an $n(np+1)$ -dimensional subvector of $\mathbf{W}_q^*(\tau)$ with variance matrix $\Upsilon_p \in \mathbb{R}^{n(np+1) \times n(np+1)}$ which is an upper-left submatrix of Υ_q .

The rest of the proof goes along the proof of Proposition 3.1 in Davis et al. (1995): The limiting distribution in (3.45) can be expressed as

$$\frac{\overline{\mathbf{W}}(\tau)^\top \Gamma \overline{\mathbf{W}}(\tau)}{\tau} + \frac{\|\overline{\mathbf{W}}(1) - \overline{\mathbf{W}}(\tau)\|^2}{1 - \tau} - \overline{\mathbf{W}}(1)^\top \Gamma \overline{\mathbf{W}}(1), \quad (3.46)$$

where $\overline{\mathbf{W}}(\tau) = \Upsilon_q^{-\frac{1}{2}} \mathbf{W}_q^*(\tau)$ and

$$\Gamma = \Upsilon_q^{\frac{1}{2}} \cdot \begin{pmatrix} \Upsilon_p^{-\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \cdot \Upsilon_q^{\frac{1}{2}} = \mathbf{U}^\top \cdot \begin{pmatrix} \mathbf{I}_{n(np+1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \cdot \mathbf{U}$$

by rotation using orthogonal matrix \mathbf{U} , since Γ is symmetric and idempotent. If we denote $\mathbf{W}(\tau) = \mathbf{U} \overline{\mathbf{W}}(\tau)$ then

$$L(\tau) = \frac{\|\mathbf{W}_1(\tau)\|^2}{\tau} + \frac{\|\mathbf{W}_1(1) - \mathbf{W}_1(\tau)\|^2 + \|\mathbf{W}_2(1) - \mathbf{W}_2(\tau)\|^2}{1 - \tau} - \|\mathbf{W}_1(1)\|^2,$$

where $\mathbf{W}(\tau) = \text{vec}(\mathbf{W}_1(\tau), \mathbf{W}_2(\tau))$ and $\mathbf{W}_1, \mathbf{W}_2$ are two independent standard Wiener processes of dimensions $n(np+1)$ and $n^2(q-p)$, respectively. \square

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

Now we present two auxiliary lemmas which will be used in the proof of Theorem 3.5 and also in Chapter 4. The proofs until the end of this section follow the similar steps as those in Davis et al. (1995).

Lemma 3.9 (a) Under H_0 and Assumptions B there exists an iid sequence $\{\mathbf{z}_t\}_{t \in \mathbb{Z}}$ on a possibly richer probability space with variance matrix $\mathbf{\Upsilon} = \mathbb{E}[\mathbf{M}_t^\top \mathbf{\Omega}^{-1} \mathbf{M}_t]$ such that

$$\mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k - \mathbf{u}_k^\top \frac{\mathbf{\Upsilon}^{-1}}{k} \mathbf{u}_k = o(1) \quad \text{a.s., as } k \rightarrow \infty,$$

where $\mathbf{u}_k := \sum_{t=p+1}^k \mathbf{z}_t$.

(b) Analogously, under H_0 and Assumptions B there exists an iid sequence $\{\tilde{\mathbf{z}}_t\}_{t \in \mathbb{Z}}$ on a possibly richer probability space with variance matrix $\mathbf{\Upsilon}$ such that

$$\tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k - \tilde{\mathbf{u}}_k^\top \frac{\mathbf{\Upsilon}^{-1}}{T-k} \tilde{\mathbf{u}}_k = o(1) \quad \text{a.s., as } (T-k) \rightarrow \infty,$$

where $\tilde{\mathbf{u}}_k := \sum_{t=k+1}^T \tilde{\mathbf{z}}_t$.

Proof: We will begin with (a). First note that matrix $\mathbf{\Upsilon}$ is correctly defined since the expected value $\mathbb{E}[\mathbf{M}_t^\top \mathbf{\Omega}^{-1} \mathbf{M}_t]$ does not depend on t due to the stationarity of the VAR(p) model and the moment conditions of $\boldsymbol{\varepsilon}_t$.

Step 1:

We will prove that $\mathbf{s}_k^\top \mathbf{\Upsilon}^{-1} \mathbf{s}_k - \mathbf{u}_k^\top \mathbf{\Upsilon}^{-1} \mathbf{u}_k = \mathcal{O}(k^{1-\lambda})$ a.s. for some $\lambda > 0$. Sequence $\{\boldsymbol{\xi}_k := \mathbf{s}_k - \mathbf{s}_{k-1} = \mathbf{M}_k^\top \mathbf{\Omega}^{-1} \boldsymbol{\varepsilon}_k\}_{k \in \mathbb{Z}}$ is an $n(np+1)$ -dimensional centered sequence. It is weakly stationary, since for all $j > 0$

$$\text{vec } \mathbb{E}[\boldsymbol{\xi}_k \boldsymbol{\xi}_{k-j}^\top] = \mathbb{E}[(\mathbf{M}_{k-j}^\top \mathbf{\Omega}^{-1} \boldsymbol{\varepsilon}_{k-j} \otimes \mathbf{M}_k^\top \mathbf{\Omega}^{-1})] \cdot \text{vec } \mathbb{E}[\boldsymbol{\varepsilon}_k] = \mathbf{0}.$$

Vectors $\boldsymbol{\xi}_k$ have finite variance matrix since

$$\begin{aligned} \text{vec } \mathbb{E}[\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top] &= \mathbb{E}[\text{vec}(\mathbf{M}_k^\top \mathbf{\Omega}^{-1} \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^\top \mathbf{\Omega}^{-1} \mathbf{M}_k)] = \\ &= \mathbb{E}[(\mathbf{M}_k^\top \mathbf{\Omega}^{-1}) \otimes (\mathbf{M}_k^\top \mathbf{\Omega}^{-1})] \cdot \text{vec}(\mathbb{E}[\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^\top]) = \\ &= \mathbb{E}[(\mathbf{M}_k^\top \mathbf{\Omega}^{-1}) \otimes (\mathbf{M}_k^\top \mathbf{\Omega}^{-1})] \cdot \text{vec}(\mathbf{\Omega}) = \\ &= \text{vec } \mathbb{E}[\mathbf{M}_k^\top \mathbf{\Omega}^{-1} \mathbf{M}_k] = \text{vec}(\mathbf{\Upsilon}). \end{aligned}$$

$\{\boldsymbol{\xi}_k\}_{k \in \mathbb{Z}}$ can be written as a measurable function $f(\mathbf{y}_k, \dots, \mathbf{y}_{k-p})$ of finitely many strong mixing random vectors and hence it is itself a strong mixing sequence with the same rate as $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ due to Theorem B.9. Since $\sup_k \mathbb{E} \|\boldsymbol{\xi}_k\|^{2+\delta} = K < \infty$, for some $\delta > 0$ and some positive K then Theorem B.10 implies that there exist a sequence of iid Gaussian random vectors $\{\mathbf{z}_t\}_{t \in \mathbb{Z}}$ on a possibly wider probability space with the variance matrix $\mathbf{\Upsilon}$ such that

$$\mathbf{s}_k - \mathbf{u}_k = \mathcal{O}(k^{\frac{1}{2}-\lambda'}) \quad \text{a.s., } k \rightarrow \infty, \quad (3.47)$$

for some $\lambda' > 0$, where $\mathbf{u}_k := \sum_{t=p+1}^k \mathbf{z}_t$. From that and the Law of the iterated logarithm for independent random vectors, it holds for the row vector $\mathbf{u}_k^\top \mathbf{\Upsilon}^{-1} = \mathcal{O}(\sqrt{k \ln \ln k})$ a.s., and hence $\mathbf{s}_k^\top \mathbf{\Upsilon}^{-1} = \mathcal{O}(\sqrt{k \ln \ln k})$ a.s., $k \rightarrow \infty$. Now

$$\begin{aligned} \mathbf{s}_k^\top \mathbf{\Upsilon}^{-1} \mathbf{s}_k - \mathbf{u}_k^\top \mathbf{\Upsilon}^{-1} \mathbf{u}_k &= \mathbf{s}_k^\top \mathbf{\Upsilon}^{-1} (\mathbf{s}_k - \mathbf{u}_k) + (\mathbf{s}_k - \mathbf{u}_k)^\top \mathbf{\Upsilon}^{-1} \mathbf{u}_k = \\ &= \mathcal{O}(\sqrt{k \ln \ln k}) \cdot \mathcal{O}(k^{\frac{1}{2}-\lambda'}) \text{ a.s.} = \mathcal{O}(k^{1-\lambda}) \text{ a.s.,} \end{aligned}$$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

$k \rightarrow \infty$, for some $\lambda > 0$.

Step 2:

To conclude, due to Step 1, it remains to show that $\mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k - \mathbf{s}_k^\top \frac{\mathbf{\Upsilon}^{-1}}{k} \mathbf{s}_k = o(1)$ a.s., $k \rightarrow \infty$:

If we show that

$$\mathbf{\Upsilon} - (k\mathbf{P}_k)^{-1} = \mathcal{O}\left(\sqrt{\frac{\ln \ln k}{k}}\right) \quad \text{a.s., } k \rightarrow \infty, \quad (3.48)$$

then

$$\mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k - \mathbf{s}_k^\top \frac{\mathbf{\Upsilon}^{-1}}{k} \mathbf{s}_k = \frac{1}{\sqrt{k}} \mathbf{s}_k^\top (k\mathbf{P}_k) \left(\mathbf{\Upsilon} - \frac{\mathbf{P}_k^{-1}}{k} \right) \mathbf{\Upsilon}^{-1} \mathbf{s}_k \frac{1}{\sqrt{k}} \rightarrow 0, \quad \text{as } k \rightarrow \infty.$$

To show (3.48), we start with the vector $\text{vec}(\mathbf{M}_t^\top \mathbf{\Omega}^{-1} \mathbf{M}_t - \mathbf{\Upsilon})$ which is centered and weakly stationary vector and its variance matrix $\mathbf{S} := \text{var } \text{vec}(\mathbf{M}_t^\top \mathbf{\Omega}^{-1} \mathbf{M}_t - \mathbf{\Upsilon})$ is not dependent on t due to Assumptions (B.1) and (B.2). Vector $\text{vec}(\mathbf{M}_t^\top \mathbf{\Omega}^{-1} \mathbf{M}_t - \mathbf{\Upsilon})$ has uniformly bounded $(2 + \eta)$ -moment, $\eta > 0$. Process $\{\text{vec}(\mathbf{M}_t^\top \mathbf{\Omega}^{-1} \mathbf{M}_t - \mathbf{\Upsilon})\}_{t \in \mathbb{Z}}$ is a strong mixing sequence with the same rate as $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ due to Assumption (B.3) and Theorem B.9. Thanks to Theorem B.10, there exists a Wiener process $\mathbf{W}_\mathbf{S}$ with variance matrix \mathbf{S} on a possibly wider probability space such that

$$\begin{aligned} \sum_{t=p+1}^k \text{vec}(\mathbf{\Upsilon} - \mathbf{M}_t^\top \mathbf{\Omega}^{-1} \mathbf{M}_t) - \mathbf{W}_\mathbf{S}(k) &= \mathcal{O}(k^{\frac{1}{2}-\lambda}) \quad \text{a.s., } k \rightarrow \infty \\ k \text{vec}(\mathbf{\Upsilon}) - \text{vec}(\mathbf{P}_k^{-1}) - \mathbf{W}_\mathbf{S}(k) &= \mathcal{O}(k^{\frac{1}{2}-\lambda}) \quad \text{a.s., } k \rightarrow \infty, \end{aligned}$$

for some $\lambda > 0$. Law of the iterated logarithm for the Wiener process, $\mathbf{W}_\mathbf{S}(k) = \mathcal{O}(\sqrt{k \ln \ln k})$ a.s., as $k \rightarrow \infty$, yields (3.48).

Proof of (b) can be done using the analogous steps as in (a) but applied on the reverse time processes

$$\tilde{\mathbf{s}}_k = \sum_{t=1}^{T-k} \mathbf{M}_{T-t+1}^\top \mathbf{\Omega}^{-1} \boldsymbol{\varepsilon}_{T-t+1}, \quad \tilde{\mathbf{P}}_k = \sum_{t=1}^{T-k} \mathbf{M}_{T-t+1}^\top \mathbf{\Omega}^{-1} \mathbf{M}_{T-t+1},$$

which also satisfy the strong mixing property. \square

Lemma 3.10 *Under H_0 and Assumptions B it holds that*

(a)

$$\lim_{\epsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P} \left[\frac{1}{a_T(n(np+1))} \left| \max_{p < k \leq T\epsilon} \Lambda_T(k) - \max_{p < k \leq T\epsilon} \left\{ \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k \right\} \right| > \delta \right] = 0,$$

(b)

$$\begin{aligned} &\lim_{\epsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P} \left[\frac{1}{a_T(n(np+1))} \left| \max_{(1-\epsilon)T \leq k < T} \Lambda_T(k) - \right. \right. \\ &\quad \left. \left. - \max_{(1-\epsilon)T \leq k < T} \left\{ \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k \right\} \right| > \delta \right] = 0. \end{aligned}$$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

Proof: From the proof of the preceding lemma we have that for $0 < \tau < 1$, $k = \lfloor T\tau \rfloor$,

$$\begin{aligned} \mathbf{s}_{\lfloor T\tau \rfloor}^\top \mathbf{P}_{\lfloor T\tau \rfloor} \mathbf{s}_{\lfloor T\tau \rfloor} &\xrightarrow[T \rightarrow \infty]{d} \frac{\|\mathbf{W}(\tau)\|^2}{\tau} \\ \tilde{\mathbf{s}}_{\lfloor T\tau \rfloor}^\top \tilde{\mathbf{P}}_{\lfloor T\tau \rfloor} \tilde{\mathbf{s}}_{\lfloor T\tau \rfloor} &\xrightarrow[T \rightarrow \infty]{d} \frac{\|\mathbf{W}(1) - \mathbf{W}(\tau)\|^2}{1 - \tau}. \end{aligned}$$

We will use the above results to establish the proof of this lemma. In particular, to establish (a), observe that

$$\begin{aligned} &\frac{\left| \max_{p < k \leq T\epsilon} \Lambda_T(k) - \max_{p < k \leq T\epsilon} \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k \right|}{a_T(n(np+1))} \leq \\ &\leq \max_{p < k \leq T\epsilon} \frac{\left| (\mathbf{s}_T - \mathbf{s}_k)^\top \tilde{\mathbf{P}}_k (\mathbf{s}_T - \mathbf{s}_k) - \mathbf{s}_T^\top \mathbf{P}_T \mathbf{s}_T \right|}{a_T(n(np+1))} \rightarrow \\ &\rightarrow \max_{t \leq \epsilon} \left| \frac{\|\mathbf{W}(1) - \mathbf{W}(t)\|^2}{1-t} - \|\mathbf{W}(1)\|^2 \right|, \quad \text{as } T \rightarrow \infty \\ &\rightarrow 0, \quad \text{as } \epsilon \rightarrow 0 \end{aligned}$$

since $a_T(n(np+1)) \rightarrow 1$, as $T \rightarrow \infty$. The proof of (b) is similar:

$$\begin{aligned} &\frac{\left| \max_{(1-\epsilon)T \leq k < T} \Lambda_T(k) - \max_{(1-\epsilon)T \leq k < T} \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k \right|}{a_T(n(np+1))} \leq \\ &\leq \max_{(1-\epsilon)T \leq k < T} \frac{\left| \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k - \mathbf{s}_T^\top \mathbf{P}_T \mathbf{s}_T \right|}{a_T(n(np+1))} \rightarrow \\ &\rightarrow \max_{(1-\epsilon) \leq t < 1} \left| \frac{\|\mathbf{W}(t)\|^2}{t} - \|\mathbf{W}(1)\|^2 \right|, \quad \text{as } T \rightarrow \infty \\ &\rightarrow 0, \quad \text{as } \epsilon \rightarrow 0. \end{aligned}$$

□

Proof of Theorem 3.5: We can proceed in the same way as in the proof of Theorem 2.2 in Davis et al. (1995). From previous Lemma 3.10 we have that as $\epsilon \rightarrow 0$,

$$\begin{aligned} \limsup_{T \rightarrow \infty} \mathbf{P} \left[\frac{\left| \max_{p < k \leq \epsilon T} \Lambda_T(k) - \max_{p < k \leq \epsilon T} \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k \right|}{a_T(n(np+1))} > \delta \right] &\rightarrow 0, \\ \limsup_{T \rightarrow \infty} \mathbf{P} \left[\frac{\left| \max_{(1-\epsilon)T < k \leq T} \Lambda_T(k) - \max_{(1-\epsilon)T < k \leq T} \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k \right|}{a_T(n(np+1))} > \delta \right] &\rightarrow 0. \end{aligned}$$

If we prove that $\forall \epsilon > 0$

$$\mathbf{P} \left[\frac{\max_{p < k \leq \epsilon T} \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k - b_T(n(np+1))}{a_T(n(np+1))} \leq x \right] \rightarrow \exp \left\{ -e^{-\frac{x}{2}} \right\}, \quad (3.49)$$

$$\mathbf{P} \left[\frac{\max_{(1-\epsilon)T < k \leq T} \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k - b_T(n(np+1))}{a_T(n(np+1))} \leq x \right] \rightarrow \exp \left\{ -e^{-\frac{x}{2}} \right\}, \quad (3.50)$$

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

as $T \rightarrow \infty$, then the rest of proof follows from the fact that for $0 < \epsilon < \frac{1}{2}$, $\max_{p < k \leq \epsilon T} \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k$ and $\max_{(1-\epsilon)T < k \leq T} \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k$ are asymptotically independent due to assumption (B.3), and thanks to the fact that

$$\frac{\max_{\epsilon T \leq k \leq (1-\epsilon)T} \Lambda_T(k) - b_T(n(np+1))}{a_T(n(np+1))} \xrightarrow[T \rightarrow \infty]{\mathbb{P}} -\infty.$$

To show (3.49) we use the result from Lemma 3.9 and

$$\begin{aligned} & \left| \max_{m \leq k \leq T} \{\mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k\} - \max_{m \leq k \leq T} \left\{ \mathbf{u}_k^\top \frac{\boldsymbol{\Upsilon}^{-1}}{k} \mathbf{u}_k \right\} \right| \leq \\ & \sup_{m \leq k} \left| \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k - \mathbf{u}_k^\top \frac{\boldsymbol{\Upsilon}^{-1}}{k} \mathbf{u}_k \right| \xrightarrow[m \rightarrow \infty]{} 0 \quad \text{a.s.}, \end{aligned} \quad (3.51)$$

where $\mathbf{u}_k = \sum_{t=p+1}^k \mathbf{z}_t$ and \mathbf{z}_t is iid centered sequence with variance matrix $\boldsymbol{\Upsilon}$. According to Lemma 2.2 of Horváth (1993b)

$$\mathbb{P} \left[\frac{\max_{p < k \leq T} \mathbf{u}_k^\top \frac{\boldsymbol{\Upsilon}^{-1}}{k} \mathbf{u}_k - b_T(n(np+1))}{a_T(n(np+1))} \leq x \right] \xrightarrow[T \rightarrow \infty]{} \exp \left\{ -e^{-\frac{x}{2}} \right\}. \quad (3.52)$$

Due to the fact that $\forall \epsilon > 0$

$$\mathbb{P} \left[\max_{p < k \leq T} \{\mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k\} = \max_{p < k \leq T\epsilon} \{\mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k\} \right] \xrightarrow[T \rightarrow \infty]{} 1,$$

(3.51) and (3.52) statement (3.49) is proven. Analogous setup can be used to prove (3.50). \square

3.5 Simulation study

The simulations have been performed in R Programming Language, see the details in R Core Team (2015). The majority of the computations has been transferred to SNĚHURKA server located at Faculty of Mathematics and Physics in Karlín which is designed for the big data and time consuming simulations. Speeding up the simulations was the key driver for using SNĚHURKA. It is worth noted here that even the current version 2.13 of R has difficulties to perform the simple least squares estimation for stationary VAR(1) process of length greater than 5 000. Hence the ability to investigate the asymptotic performance of the test statistic for larger sample size was another reason for using this server.

Throughout this section we will work with the simulations of 2-dimensional VAR(1) processes $P_1 : \mathbf{y}_t = \boldsymbol{\Phi}^{(1)} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$ and $P_2 : \mathbf{y}_t = \boldsymbol{\Phi}^{(2)} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$, with

$$\boldsymbol{\Phi}^{(1)} = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.1 \end{pmatrix}, \quad \boldsymbol{\Phi}^{(2)} = \begin{pmatrix} 0.8 & 0.3 \\ 0.1 & 0.7 \end{pmatrix}.$$

The error process will be either

[E1] iid $\mathcal{N}_2(0, \boldsymbol{\Omega})$, where

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix},$$

[E2] iid t_ν with $\nu = 5$ degrees of freedom and variance matrix $\frac{\nu}{\nu-2}\mathbf{\Omega}$.

[E3] independent $\mathcal{N}_2(0, \mathbf{\Omega}_t)$, where

$$\mathbf{\Omega}_t = \begin{pmatrix} 3 \cdot |\cos(R_t)| & 0 \\ 0 & 3 \cdot |\cos(R_t)| \end{pmatrix},$$

where R_t has uniform distribution over $(0,1)$.

[E4] VAR(1) model

$$\boldsymbol{\varepsilon}_t = \mathbf{A}(\rho)\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\eta}_t \quad \text{with} \quad \mathbf{A}(\rho) = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix}$$

with $\rho = 0.5$ and $\boldsymbol{\eta}_t$ iid $\mathcal{N}_2(0, \mathbf{I}_2)$.

The first two conditions on errors [E1] and [E2] satisfy Assumptions A, B and C. Error sequence [E3] violates the condition of constant unconditional variance matrix reported in all assumptions. [E4] violates the martingale difference assumption in Assumptions A however it has constant conditional and unconditional variance. As regards Assumptions B and C, [E4] violates conditions (B.2) and (C.4) since $\boldsymbol{\varepsilon}$ is serially correlated. In case of [E4], the least squares estimator of $\boldsymbol{\beta}$ is not consistent.

Characteristic polynomials of both processes P_1 and P_2 have roots outside the unit circle, but the roots of the polynomial $\det\{\mathbf{I}_2 - \boldsymbol{\Phi}^{(2)}z\}$ of process P_2 are close to the unit circle. Processes P_1 and P_2 were generated 1 000 times.

3.5.1 Known time point

First, we consider the situation when a time point is known a priori and let us consider the length of the time series $T = 200$. For each process P_i , $i = 1, 2$, the empirical distribution function (=EDF) of $Q_T(\tau)$ is plotted for 3 different values of τ together with the asymptotic distribution functions (=ADF) of the corresponding limiting processes

$$\frac{\|\mathbf{W}(\tau) - \tau\mathbf{W}(1)\|^2}{\tau(1-\tau)}. \tag{3.53}$$

The result is reported in Figure 3.1. We can see good convergence results in case of process P_1 for all types of errors and locations of the break point τ . Further experiments not depicted here show that if τ is even closer to the beginning or end of the data sample the simulations get worse. As regards P_2 , Figure 3.1 illustrates fast convergence under [E1], [E2] and [E3] even for moderate length of the time series. Hence, unconditional heteroscedasticity in [E3] gives reasonable results for both P_1 and P_2 , which are comparable to situations under [E1] and [E2]. The serial correlation in the error term employed in [E4] gives promising results with somewhat worse convergence in case of P_2 . Further, under P_2 , [E4], time point τ starts to influence the results.

However, if the length of the time series starts to increase as plotted in Figure 3.2 we can see that the approximation gets better even for [E3] and [E4]. Situation is plotted for $\tau = \frac{1}{4}$.

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

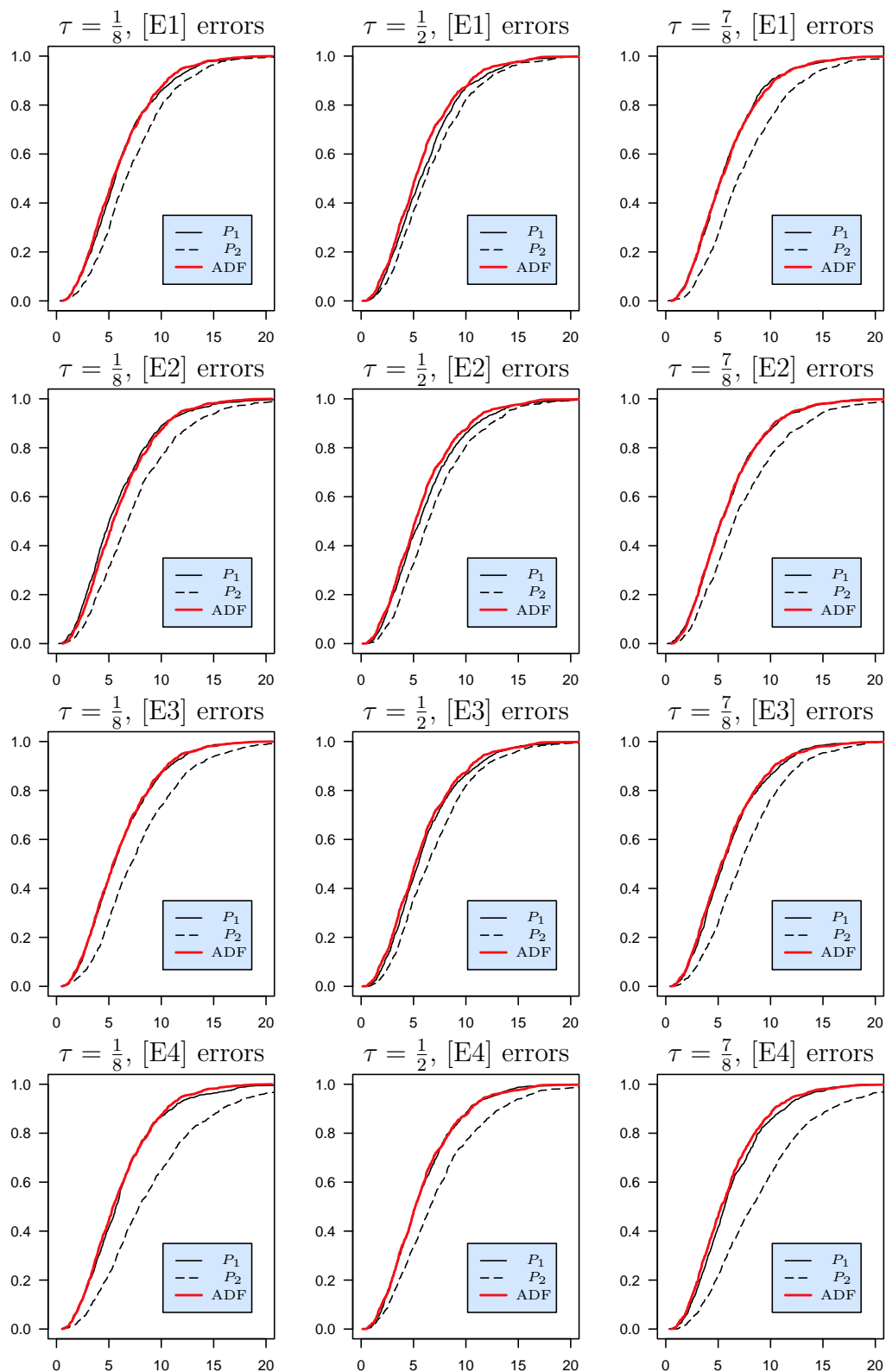


Figure 3.1: Known change point. EDF of $Q_T(\tau)$ and ADF of (3.53) for various τ . Error terms are [E1], [E2], [E3] and [E4].

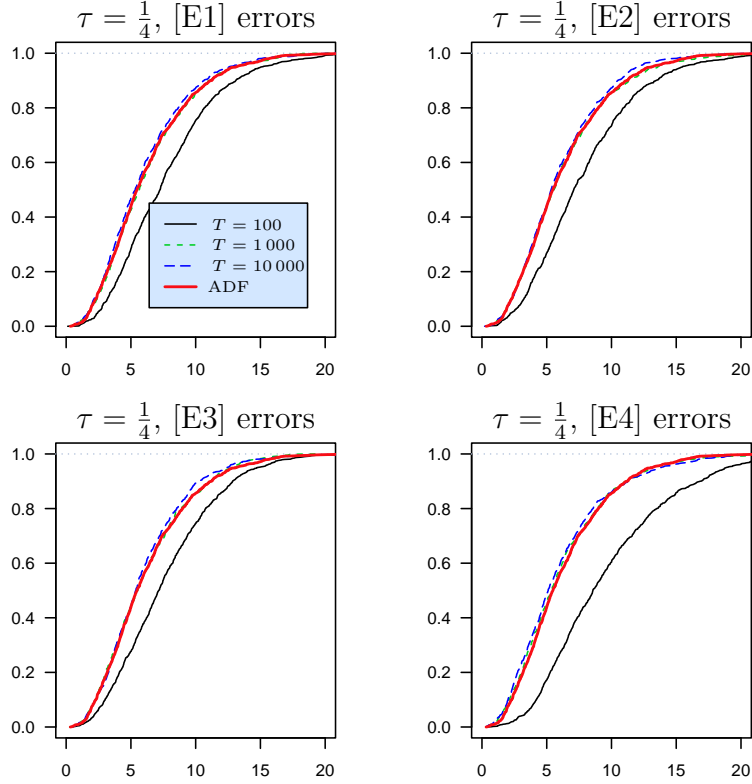


Figure 3.2: Known change point. EDF of $Q_T(\tau)$, $\tau = \frac{1}{4}$, and ADF for process P_2 for different lengths T and different white noise sequences. The legend for all the graphs is the same as in the top left figure.

3.5.2 Unknown time point

We study the asymptotic behaviour of the statistics for the same processes and errors as in the previous subsection. In particular, the simulation of

$$\sup_{t_1 < k < t_2} \Lambda_T(k), \quad t_1 = \eta T, \quad t_2 = (1 - \eta)T, \quad \eta = 0.05 \quad (3.54)$$

ran 1 000 times and the critical values were computed from the empirical distribution of the simulated statistic (3.54). The results were compared with the simulated critical values of the limiting process

$$\sup_{\eta < \tau < 1 - \eta} \left\{ \frac{\|\mathbf{W}(\tau) - \tau \mathbf{W}(1)\|^2}{\tau(1 - \tau)} \right\} \quad (3.55)$$

which is denoted as ‘‘Approximation 1’’ in Table 3.1. These values were also compared with the approximate critical values derived in James et al. (1987) (denoted as ‘‘Approximation 2’’):

$$\begin{aligned} & \mathbb{P} \left[\max_{\eta < \tau < 1 - \eta} \left\{ \frac{\|\mathbf{W}(\tau) - \tau \mathbf{W}(1)\|^2}{\tau(1 - \tau)} \right\} > c \right] \\ & \text{is approximately equal to} \\ & c^{\frac{d}{2}} \cdot \frac{e^{-\frac{c}{2}}}{2^{\frac{d-2}{2}}} \cdot \Gamma\left(\frac{d}{2}\right) \cdot \left(\left(1 - \frac{d}{c}\right) \cdot \ln\left(\frac{1}{\eta} - 1\right) + \frac{2}{c} \right), \end{aligned} \quad (3.56)$$

for large values of c , where $d = n(np + 1)$ is the dimension of the Brownian bridge process, i.e., $d = 6$ in our case. Simulations of (3.55) were also done

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

in Bai and Perron (1998), p. 58, for case $\eta = 0.05$ and several dimensions and multiple changes. Their results are reported as “Approximation 3” in Table 3.1. According to Table 3.1, both approximations appear to yield similar results even if Approximation 2 was derived for large values of c .

Table 3.1: Critical values of (3.54) for different T and approximations based on (3.55) and (3.56). Errors [E1] considered.

T	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
	P_1	P_2	P_1	P_2	P_1	P_2
100	18.04	21.77	20.04	23.80	23.29	27.85
1 000	20.10	21.13	21.72	23.47	26.27	27.54
10 000	19.31	19.68	21.59	22.06	26.06	25.99
Approximation 1	19.51		21.62		26.52	
Approximation 2	19.94		21.98		26.40	
Approximation 3	19.38		21.59		25.95	

The empirical distribution function of statistic (3.54) for different values of T was compared with the ADF of the limiting process in (3.55). The results are depicted in Figure 3.3. Generally, the quality of approximation increases with increasing length T . In case of P_1 there is a very fast convergence under all type of error term processes. This is also true in case P_2 , for [E1]–[E3]. In case of P_2 , for [E4], the convergence is somewhat slower.

3.5.3 Change in the lag

The next simulation study considers the possibility of a lag change after the break point. We compare the approximations for processes P_1 and P_2 with errors [E1] – [E3] under H_0 . Under [E4] serial correlation in the error term violates the results substantially. The trimming has been done in the same way as in (3.54), hence 5 % of observations from both sides of the time series were cut. The results are reported in Figure 3.4.

Figure 3.4 illustrates that only in case of process P_2 with errors [E2], the convergence is somewhat slower compared to other scenarios. The empirical quantiles of the test statistic on the left-hand side of (3.8) for different levels α for both processes P_1 and P_2 with errors [E1] can be seen in Table 3.2 together with the simulated quantiles of the asymptotic distribution on the right-hand side of (3.8).

3.5.4 Approximation by Gumbel distribution

The simulation setting is the same as in the previous subsections. The empirical results based on Theorem 3.5 are illustrated in Figure 3.5. The limiting distribution tends to be smaller than the empirical one for all cases. The speed of convergence is much slower even for the error term [E1]. Generally slower convergence to the Gumbel distribution was independently confirmed also by Hall (1979), Horváth (1993b), or Aue and Horváth (2012) among others. More reasonable results are achieved for errors [E3] than for errors [E2].

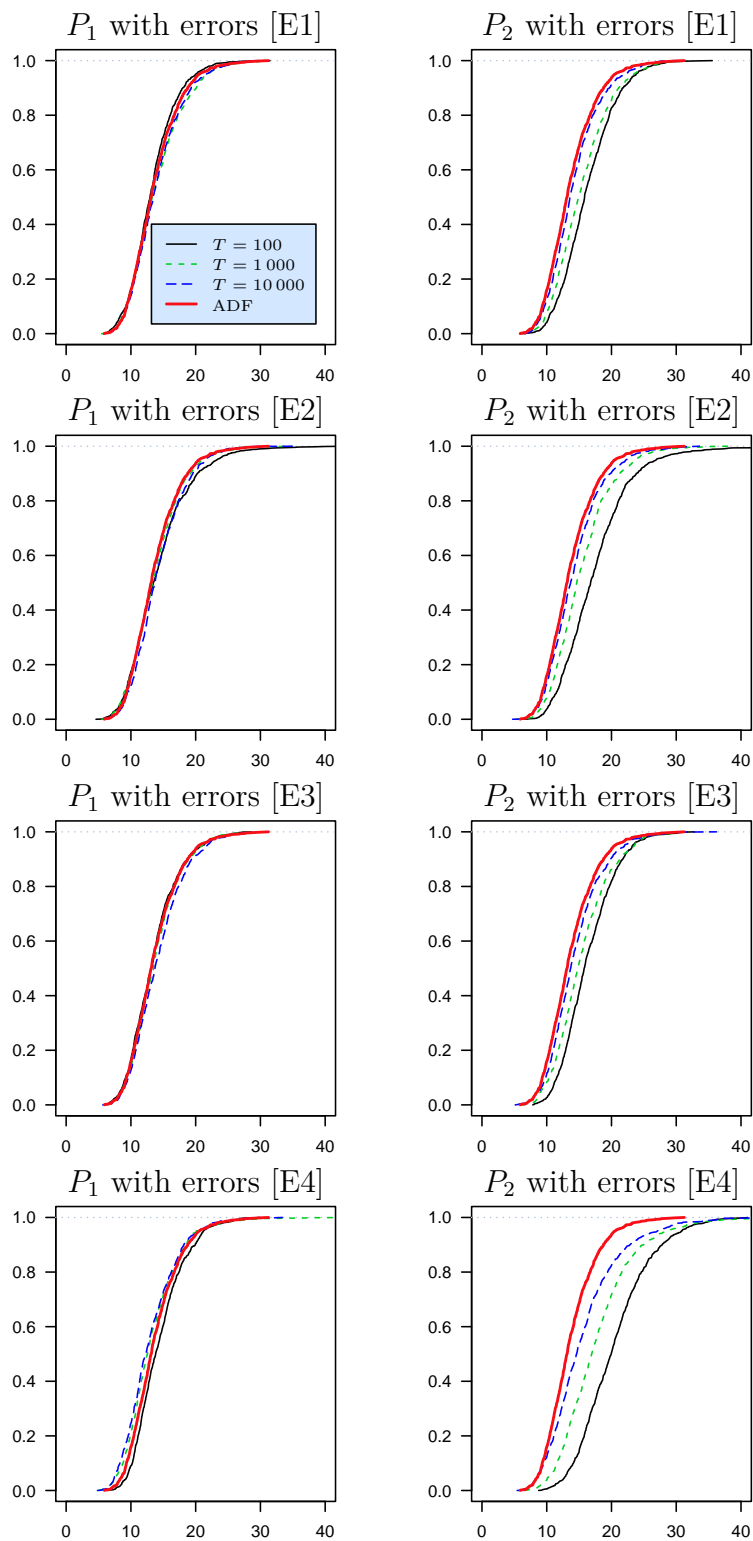


Figure 3.3: Unknown change point. EDF of (3.54) for different T compared with ADF of (3.55). The legend for all the graphs is the same as in the top left figure.

3.5.5 Empirical power

Let us briefly discuss the empirical power of the proposed tests. As regards proving asymptotic consistency, it is rather complicated to prove it in case of LR test since many terms of the likelihood ratio should be taken into consideration.

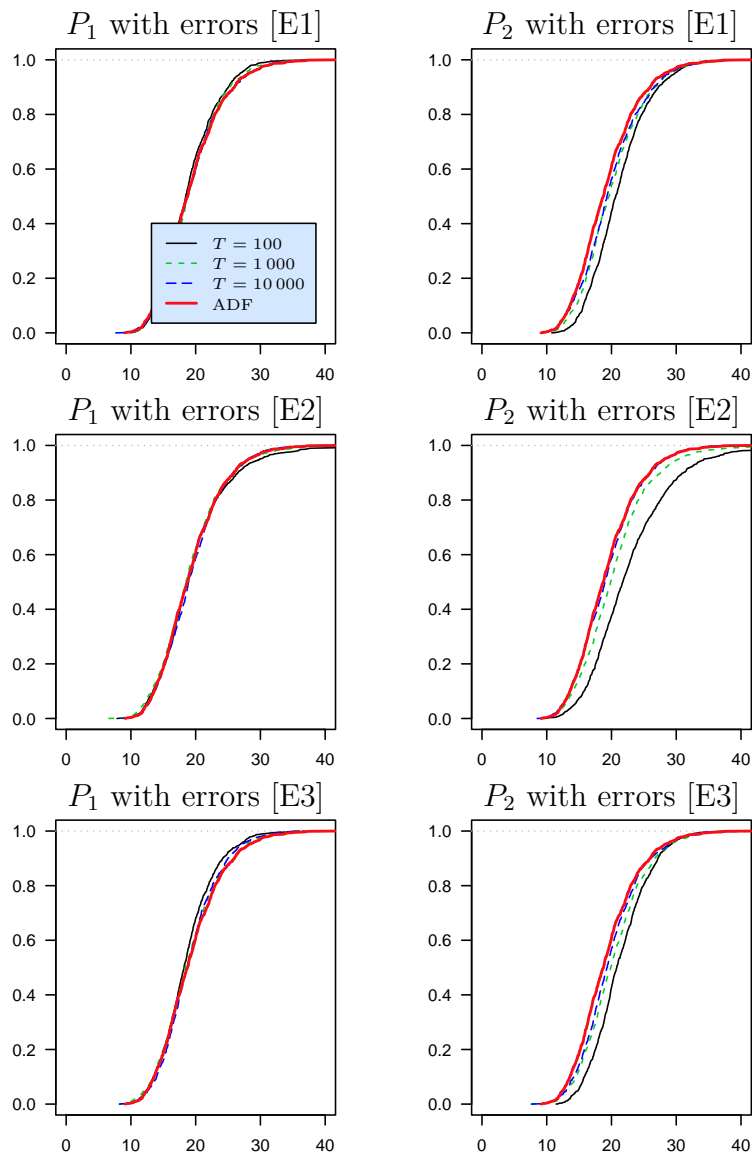


Figure 3.4: Empirical and asymptotic distribution functions of approximation (3.8) for processes P_1 and P_2 with different T and various error terms.

Table 3.2: Critical values of Q_T^\dagger for different T . Errors [E1] were considered.

T	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
	P_1	P_2	P_1	P_2	P_1	P_2
100	25.04	27.24	26.85	29.76	30.47	33.95
1000	25.28	26.75	27.55	28.73	32.96	34.72
10000	26.02	26.58	28.01	28.86	32.38	33.61
Asympt. distr.	25.98		28.15		33.79	

Hence at least simulation results which signalize promising results under alternative hypothesis will be reported next. All simulations in this subsection have been done for SCENARIO 1 and process P_1 of length $T = 1000$ with errors being iid [E1] and heteroscedastic [E3]. The empirical sizes for process P_2 lie somewhat

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

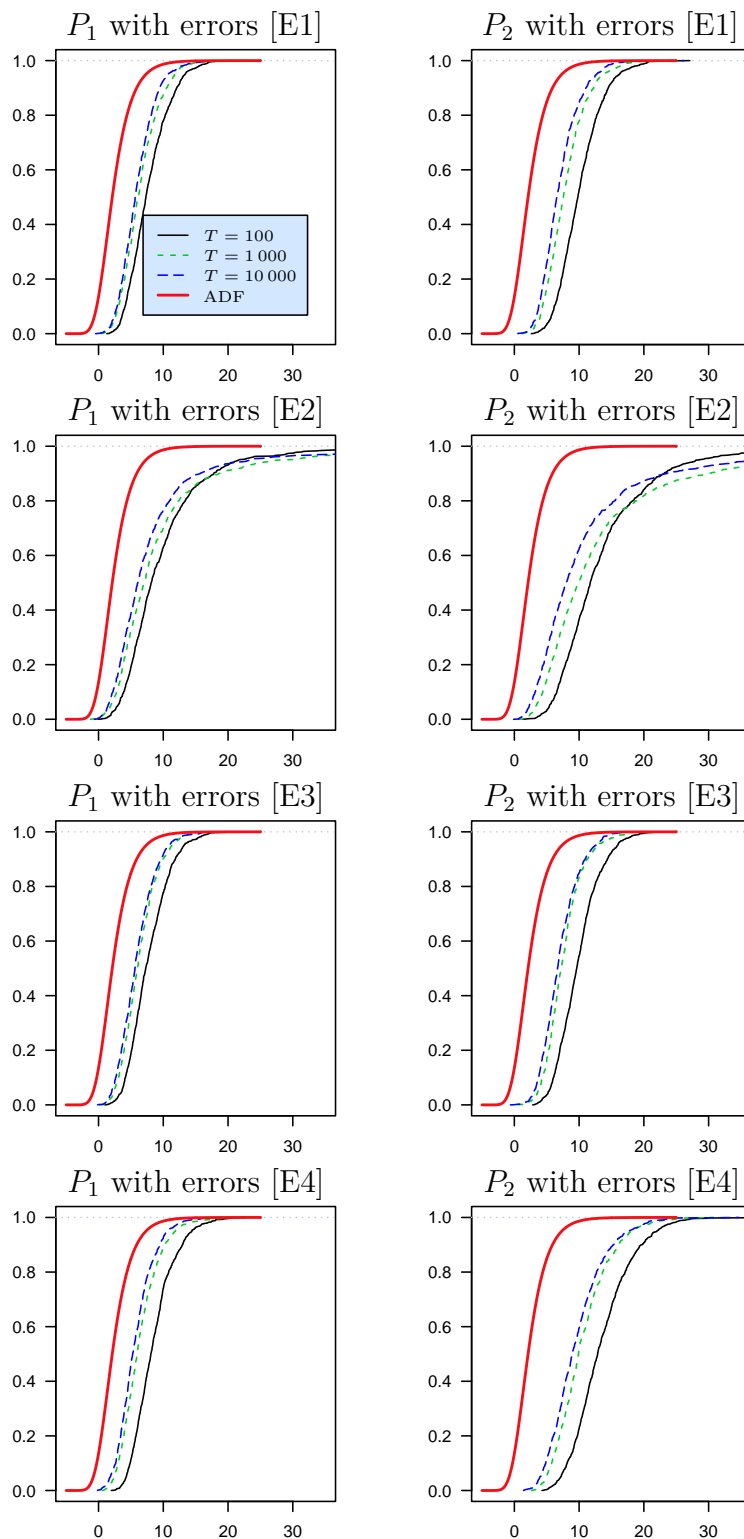


Figure 3.5: Empirical and asymptotic distribution functions of approximation (3.9) for processes $[P_1]$ and $[P_2]$ with different T . The legend in the top left plot applies to other cases as well.

lower compared to P_1 for both cases [E1] and [E3]. The change point occurs exactly in the middle of the simulated series at $k = \frac{T}{2} = 500$. The test level will be 0.05 in all cases and the critical value is set with respect to the situation when all autoregressive parameters and intercept are subject of a change. This situation

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

occurs most often in practical applications since usually an analyst does not know which parameters are subject of a change. Of course, having apriori information and assuming only intercept change will lead to the increase of power of the test.

First, we will discuss change in \mathbf{c} from $(0, 0)^\top$ to $\tilde{\mathbf{c}} = (\tilde{c}_1, \tilde{c}_2)^\top$ with other parameters being unchanged. Table 3.3 shows the difference in terms of empirical power of the test statistic under [E1] and [E3]. We can see that the heteroscedasticity in the error term causes a substantial decrease in terms of power of the test for small changes of intercept. Approximately, test reaches empirical power one for more than twice as big change under heteroscedasticity than under iid errors. The biggest decrease of power as well as the lowest distance for which the test firstly have empirical power one are highlighted in blue in Table 3.3.

Table 3.3: Empirical sizes of the test.

\tilde{c}_1	\tilde{c}_2	$\sqrt{\tilde{c}_1^2 + \tilde{c}_2^2}$	power [E1]	power [E3]	% decr.
0.04	-0.08	0.09	0.09	0.05	-51
0.08	-0.05	0.10	0.11	0.05	-53
0.00	0.13	0.13	0.17	0.06	-65
0.05	-0.16	0.17	0.32	0.07	-80
0.17	0.10	0.20	0.30	0.10	-67
0.23	0.01	0.23	0.56	0.09	-84
0.13	-0.23	0.27	0.85	0.13	-85
-0.33	0.08	0.34	0.96	0.17	-82
-0.06	-0.35	0.36	0.95	0.18	-81
-0.09	0.38	0.39	0.99	0.23	-76
-0.41	-0.11	0.43	0.99	0.29	-71
-0.35	-0.35	0.49	1.00	0.39	-61
-0.45	-0.23	0.50	1.00	0.42	-58
0.42	0.35	0.55	1.00	0.49	-51
-0.45	-0.41	0.60	1.00	0.60	-40
0.69	0.01	0.69	1.00	0.72	-28
0.47	-0.59	0.75	1.00	0.81	-19
-0.60	-0.48	0.77	1.00	0.82	-18
-0.06	0.82	0.82	1.00	0.89	-11
0.81	0.33	0.88	1.00	0.92	-8
0.72	-0.65	0.97	1.00	0.96	-4
0.39	0.91	0.99	1.00	0.97	-3
0.65	-0.93	1.14	1.00	0.99	-1
-0.92	0.84	1.24	1.00	1.00	0

The plot in Figure 3.6 represents the power function under the alternative hypothesis when the parameter \mathbf{c} changes from the origin to the different vector. On x axis there is the Euclidean distance of new vector $\tilde{\mathbf{c}}$ from origin.

Empirical sizes of the test when autoregression parameters change are depicted in Table 3.4. Simulation of the alternative has been performed such that the VAR(1) process was again stationary after the change. The difference between the value of element of the autoregression matrix after and before the change is marked in blue. The darker the color the bigger the change.

Simulation results show that the heteroscedasticity influences the result much

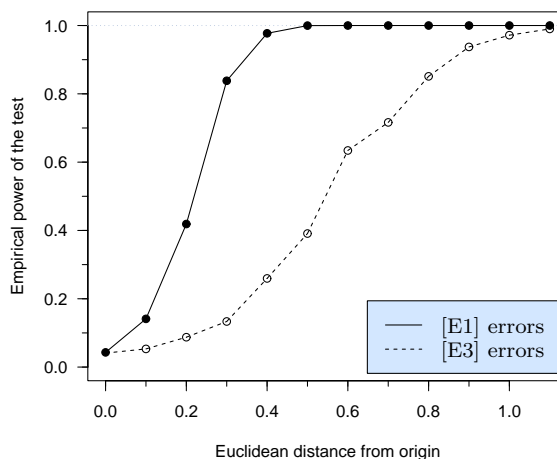


Figure 3.6: Empirical power simulated for process $[P_1]$ of length $T = 1000$ with $[E1]$ and $[E3]$ error term sequence.

Table 3.4: Empirical sizes of the test.

$\tilde{\varphi}_{11}$	$\tilde{\varphi}_{12}$	$\tilde{\varphi}_{21}$	$\tilde{\varphi}_{22}$	power	power
				[E1]	[E3]
0.5	0.2	0.2	0.1	0.05	0.04
0.5	0.2	0.2	0.0	0.09	0.12
0.5	0.2	0.2	0.2	0.11	0.09
0.5	0.0	0.2	0.1	0.48	0.42
0.5	0.2	0.2	0.3	0.48	0.45
0.3	0.2	0.2	0.1	0.57	0.52
0.5	0.2	0.0	0.1	0.62	0.57
0.5	0.2	0.2	0.4	0.94	0.93
0.2	0.2	0.2	0.1	0.95	0.92
0.5	0.2	0.2	0.5	1.00	1.00
0.4	0.2	0.2	0.0	0.22	0.21
0.5	0.1	0.0	0.1	0.69	0.67
0.3	0.2	0.2	0.3	0.92	0.90
0.5	0.4	0.4	0.1	0.93	0.94
0.2	0.2	0.2	0.0	0.95	0.96
0.5	0.3	0.5	0.1	0.99	0.99

less in case of changes in the autoregression than in case of intercept change.

3.6 Application

We applied the results to a time series of 888 monthly US stock log-returns of IBM and S&P collected between January 1926 and December 1999. Even if ARCH or GARCH models are more frequent for financial data than autoregressions, a VAR is shown to be one of the possible models, see Tsay (2010).

Justification for using a multivariate model in this example instead of analyz-

3. LR TEST WHERE VARIANCE OF ERRORS IS UNCHANGED

ing each data set separately may be seen in the scatterplot in Figure 3.7. This figure indicates potential dependence between these two series where the sample correlation is 0.64.

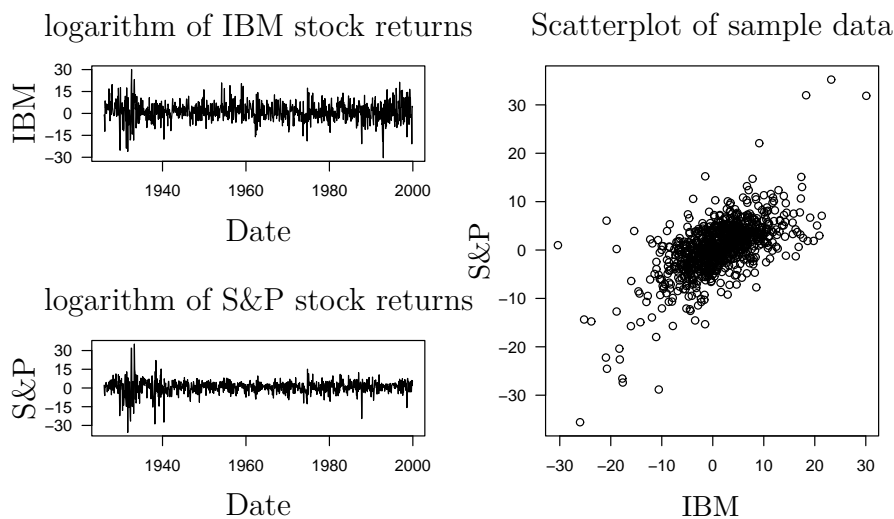


Figure 3.7: Visualization of logarithm of IBM and S&P stock returns.

On the basis of the AIC criterion (see Hamilton (1994), Chapter 4.5), we considered a VAR(5) model. The performance of the test statistic $\Lambda_T(k)$ with changing k is plotted in Figure 3.8 for the time period from September 1929 to March 1996.

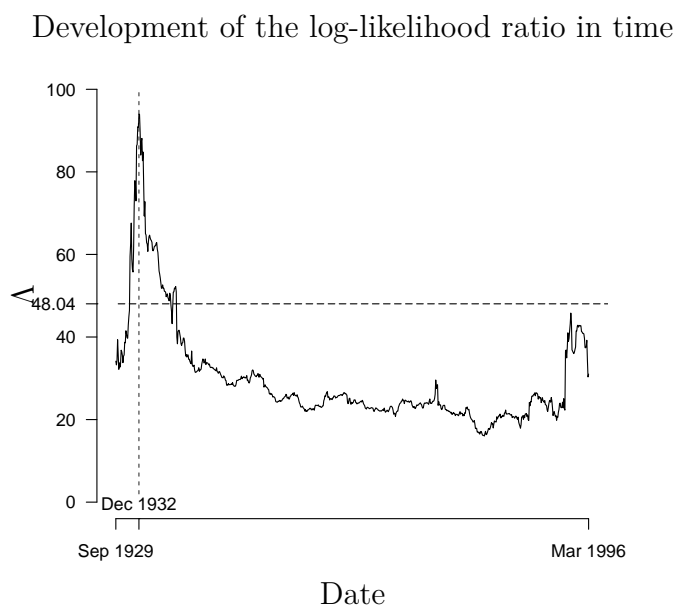


Figure 3.8: $\Lambda_T(k)$ for logarithm of IBM and S&P returns.

It is clear that the maximum of the likelihood ratio is achieved in December 1932. At that time, the Great Depression affected the whole US as a result of the fear following the market crash in the autumn of 1929. The critical value for $\alpha = 0.05$ (horizontal dashed line) is obtained from repeated simulation of (3.55),

where $\mathbf{W}(\cdot)$ in that formula is a standard $n(np + 1) = 22$ -dimensional Wiener process.

According to Tsay (2010), a bivariate GARCH(1,1) model is also considered as one of the possible models for our data. The existence of a change point for such models can be tested using the CUSUM testing procedure published in Kokoszka and Leipus (2000).

3.7 Chapter Summary

Testing procedures for changes in autoregressive parameters in a VAR(p) model have been studied under various sets of assumptions on the error term. Under the null hypothesis, we have found the asymptotic properties of the derived likelihood ratio test statistics in cases of known and unknown change and for the case that the lag of the model can increase. We have shown that under Assumptions B and SCENARIO 1 one can find the approximation of quasi-likelihood test statistic by Gumbel distribution which was previously done only in case of univariate models. The extension to SCENARIO 2 can be part of the future study however we think based on what we have done that the extension can be done with a relatively little effort based on what has already been proven in Davis et al. (1995). As regards other conditions A or C the extension could be done provided that we would have a certain speed in the FCLTs.

The quality of the convergence has been illustrated on many simulation examples. Even situations which do not fulfil some of the conditions have been considered. Surprisingly, heteroscedasticity in variance of the error term does not influence much the quality of approximation of the test statistics by asymptotic distribution and this result can be a motivation for further relaxation of the moment assumptions in the theoretical part. However we have found out in our simulation examples that heteroscedasticity has in overall a negative impact on the power of the test under the alternative hypothesis. The test loses the power especially under the intercept change; change in the autoregression structure has a relatively small impact. Simulation studies have revealed that much more important assumption is the uncorrelatedness of the error term. Serial correlation of errors in VAR(p) model can severely violate the simulation results due to the inconsistency of the least squares estimators. One possible solution could be the derivation of likelihood-ratio test specially for the serial correlation of the error term. Simulation results for the Darling-Erdős-type test have revealed that the convergence is substantially slower compared to the approximations by functionals of the Wiener process.

In this chapter we have focused on cases where the variance matrix of the error term has not undergone a change. However, in a variety of applications it is useful to detect a change, both in the autoregressive matrix and the variance of the error term. The Darling-Erdős-type test for such situation will be treated in Chapter 4.

4. Darling-Erdős test for changes in variance

In this chapter we will consider SCENARIO 3 and assume that both variance of the error term and autoregression parameters might change. We will discuss the possibility to approximate the quasi-likelihood ratio test statistic by the Gumbel distribution under the null hypothesis similarly as in the univariate case, see for instance Davis et al. (1995). Unlike in the latter article, however, we will need a more moment-restrictive modification of Assumptions B in order to prove the asymptotic results of the test statistic. The reason is that the test statistic becomes more complicated in higher dimensions.

We will show that unlike in univariate AR models, classical quasi-likelihood ratio test statistic under the null hypothesis and SCENARIO 3 does not converge to the Gumbel distribution. This seems to be a quite unexpected result since in case of SCENARIO 1 there exists a Gumbel approximation of the quasi-likelihood ratio test statistic under H_0 for both univariate and multivariate VAR(p) process under similar sets of conditions.

Section 4.1 describes the test statistic for SCENARIO 3. Section 4.2 contains the Taylor expansion of log-likelihood ratio which is the key tool to reveal its asymptotic properties. Comparison of the univariate and multivariate setups clarifies the reasons why there exists the Gumbel distribution as a proxy for the univariate quasi-likelihood approach and does not exist in the multivariate case. The modification of the test statistic will be proposed for which we are able to find the asymptotic Gumbel distribution under H_0 . Section 4.4 provides auxiliary lemmas with proofs and the proof of the main theorem as well. Simulation study in Section 4.5 documents our findings. The reasons of somewhat slower convergence will also be discussed. Most of the theoretical part of this chapter has been published in Dvořák (2015).

4.1 Test derivation

We will again consider VAR(p) process (2.1) and the notation introduced in Chapter 2. The quasi-likelihood ratio in SCENARIO 3 based on the multivariate Gaussian distribution is defined as

$$LR' := \frac{\sup_{\beta, \tilde{\beta}, \Omega, \tilde{\Omega}} \left\{ \prod_{t=p+1}^k f_t(\beta, \Omega) \cdot \prod_{t=k+1}^T f_t(\tilde{\beta}, \tilde{\Omega}) \right\}}{\sup_{\beta, \Omega} \left\{ \prod_{t=p+1}^T f_t(\beta, \Omega) \right\}},$$

where

$$f_t(\beta, \Omega) = (2\pi)^{-\frac{n}{2}} |\Omega|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{M}_t \beta)^\top \Omega^{-1} (\mathbf{y}_t - \mathbf{M}_t \beta) \right\}$$

is again a multivariate Gaussian density.

If we carry out $2 \cdot \ln(LR')$ transformation and plug in the least squares estimators $\hat{\beta}_T, \hat{\beta}_k, \tilde{\hat{\beta}}_k, \hat{\Omega}_T, \hat{\Omega}_k$ and $\tilde{\hat{\Omega}}_k$ as in Chapter 3, we can define the test statistic

of the form $\Lambda'_T := \max_{p < k \leq T} \{\Lambda'_T(k)\}$, where

$$\Lambda'_T(k) = (T-p) \ln |\widehat{\boldsymbol{\Omega}}_T| - (k-p) \ln |\widehat{\boldsymbol{\Omega}}_k| - (T-k) \ln |\widehat{\widehat{\boldsymbol{\Omega}}}_k|, \quad (4.1)$$

with

$$\widehat{\boldsymbol{\Omega}}_T := \frac{\mathbf{Q}_T}{T-p}, \quad \widehat{\boldsymbol{\Omega}}_k := \frac{\mathbf{Q}_k}{k-p}, \quad \widehat{\widehat{\boldsymbol{\Omega}}}_k := \frac{\widetilde{\mathbf{Q}}_k}{T-k}$$

and

$$\begin{aligned} \mathbf{Q}_T &:= \sum_{t=p+1}^T (\mathbf{y}_t - \mathbf{M}_t \widehat{\boldsymbol{\beta}}_T) (\mathbf{y}_t - \mathbf{M}_t \widehat{\boldsymbol{\beta}}_T)^\top, \\ \mathbf{Q}_k &:= \sum_{t=p+1}^k (\mathbf{y}_t - \mathbf{M}_t \widehat{\boldsymbol{\beta}}_k) (\mathbf{y}_t - \mathbf{M}_t \widehat{\boldsymbol{\beta}}_k)^\top, \\ \widetilde{\mathbf{Q}}_k &:= \sum_{t=k+1}^T (\mathbf{y}_t - \mathbf{M}_t \widehat{\widehat{\boldsymbol{\beta}}}_k) (\mathbf{y}_t - \mathbf{M}_t \widehat{\widehat{\boldsymbol{\beta}}}_k)^\top. \end{aligned}$$

4.2 Darling-Erdős type approximation

In order to study the asymptotic properties of Λ'_T it is better to consider the second-order Taylor expansion of the logarithmic functions appearing in (4.1). This decomposition reveals the core reason why the multivariate quasi-likelihood test statistic does not converge to the Gumbel distribution.

Considering Assumptions B, term $\ln |\widehat{\boldsymbol{\Omega}}_T|$ can be expanded around the true value $\ln |\boldsymbol{\Omega}|$ such that

$$\ln |\widehat{\boldsymbol{\Omega}}_T| = \ln |\boldsymbol{\Omega}| + \text{tr} \left\{ \boldsymbol{\Omega}^{-1} (\widehat{\boldsymbol{\Omega}}_T - \boldsymbol{\Omega}) \right\} - \frac{1}{2} \text{tr} \left\{ \left(\boldsymbol{\Omega}^{-1} (\widehat{\boldsymbol{\Omega}}_T - \boldsymbol{\Omega}) \right)^2 \right\} + o_P \left(\frac{1}{T} \right),$$

as $T \rightarrow \infty$, since due to the central limit theorem for strong mixing sequences, $\widehat{\boldsymbol{\Omega}}_T - \boldsymbol{\Omega} = \mathcal{O}_P(T^{-\frac{1}{2}})$, as $T \rightarrow \infty$.

Let us consider conditions

$$\frac{1}{T} \sum_{t=p+1}^{\lfloor T\tau \rfloor} \mathbf{M}_t^\top \mathbf{M}_t \xrightarrow[T \rightarrow \infty]{P} \tau \mathbf{Q} \quad \text{uniformly for all } \tau \in [(p+1)/T, 1], \quad (4.2)$$

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{\lfloor T\tau \rfloor} \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \xrightarrow[T \rightarrow \infty]{d} \boldsymbol{\Upsilon}^{\frac{1}{2}} \mathbf{W}(\tau), \quad \tau \in [(p+1)/T, 1], \quad (4.3)$$

where \mathbf{Q} is some positive-definite matrix, $\boldsymbol{\Upsilon} := \mathbf{E}[\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t]$ and \mathbf{W} is a standard Wiener process with independent components. As shown in Chapter 3, under Assumptions B, statements (4.2) and (4.3) hold due to FCLT B.10 for strong mixing sequences. In light of these uniform conditions (4.2) and (4.3), we can apply Taylor expansion on all terms in (4.1) to show that

$$\begin{aligned} \Lambda'_T(k) &= (T-p) \text{tr} \left\{ \boldsymbol{\Omega}^{-1} \widehat{\boldsymbol{\Omega}}_T \right\} - (k-p) \text{tr} \left\{ \boldsymbol{\Omega}^{-1} \widehat{\boldsymbol{\Omega}}_k \right\} - (T-k) \text{tr} \left\{ \boldsymbol{\Omega}^{-1} \widehat{\widehat{\boldsymbol{\Omega}}}_k \right\} - \\ &\quad - \frac{T-p}{2} \text{tr} \left\{ \left(\boldsymbol{\Omega}^{-1} \widehat{\boldsymbol{\Omega}}_T - \mathbf{I}_n \right)^2 \right\} + \frac{k-p}{2} \text{tr} \left\{ \left(\boldsymbol{\Omega}^{-1} \widehat{\boldsymbol{\Omega}}_k - \mathbf{I}_n \right)^2 \right\} + \\ &\quad + \frac{T-k}{2} \text{tr} \left\{ \left(\boldsymbol{\Omega}^{-1} \widehat{\widehat{\boldsymbol{\Omega}}}_k - \mathbf{I}_n \right)^2 \right\} + o_P(1), \quad p < k \leq T. \end{aligned} \quad (4.4)$$

Summing up the first three addends in expression (4.4) gives

$$\begin{aligned}
 & \text{tr}\{\Omega^{-1}\mathbf{Q}_T\} - \text{tr}\{\Omega^{-1}\mathbf{Q}_k\} - \text{tr}\{\Omega^{-1}\tilde{\mathbf{Q}}_k\} = \\
 &= \sum_{t=p+1}^T (\mathbf{y}_t - \mathbf{M}_t\hat{\boldsymbol{\beta}}_T)^\top \Omega^{-1}(\mathbf{y}_t - \mathbf{M}_t\hat{\boldsymbol{\beta}}_T) - \\
 & \quad - \sum_{t=p+1}^k (\mathbf{y}_t - \mathbf{M}_t\hat{\boldsymbol{\beta}}_k)^\top \Omega^{-1}(\mathbf{y}_t - \mathbf{M}_t\hat{\boldsymbol{\beta}}_k) - \\
 & \quad - \sum_{t=k+1}^T (\mathbf{y}_t - \mathbf{M}_t\hat{\boldsymbol{\beta}}_k)^\top \Omega^{-1}(\mathbf{y}_t - \mathbf{M}_t\hat{\boldsymbol{\beta}}_k) =: Q_T - Q_k - \tilde{Q}_k.
 \end{aligned}$$

If we denote as before

$$\begin{aligned}
 \mathbf{s}_k &:= \sum_{t=p+1}^k \mathbf{M}_t^\top \Omega^{-1} \boldsymbol{\varepsilon}_t, & \tilde{\mathbf{s}}_k &:= \sum_{t=k+1}^T \mathbf{M}_t^\top \Omega^{-1} \boldsymbol{\varepsilon}_t, \\
 \mathbf{P}_k &:= \left(\sum_{t=p+1}^k \mathbf{M}_t^\top \Omega^{-1} \mathbf{M}_t \right)^{-1}, & \tilde{\mathbf{P}}_k &:= \left(\sum_{t=k+1}^T \mathbf{M}_t^\top \Omega^{-1} \mathbf{M}_t \right)^{-1}, \\
 \mathbf{e}_k &:= \text{vec}(\boldsymbol{\varepsilon}_{p+1}, \dots, \boldsymbol{\varepsilon}_k), & \tilde{\mathbf{e}}_k &:= \text{vec}(\boldsymbol{\varepsilon}_{k+1}, \dots, \boldsymbol{\varepsilon}_T),
 \end{aligned}$$

then it holds that

$$\begin{aligned}
 Q_T &= \mathbf{e}_T^\top (\mathbf{I}_{T-p} \otimes \Omega^{-1}) \mathbf{e}_T - \mathbf{s}_T^\top \mathbf{P}_T \mathbf{s}_T, & Q_k &= \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k - \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k, \\
 \tilde{Q}_k &= \tilde{\mathbf{e}}_k^\top (\mathbf{I}_{T-k} \otimes \Omega^{-1}) \tilde{\mathbf{e}}_k - \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k.
 \end{aligned} \tag{4.5}$$

Hence, omitting asymptotically negligible terms in (4.4) the Taylor expansion of the quasi-likelihood ratio can be approximated under H_0 by

$$\Lambda_T(k) + \underbrace{\frac{k-p}{2} \text{tr}\left\{(\Omega^{-1}\hat{\boldsymbol{\Omega}}_k - \mathbf{I}_n)^2\right\}}_{\text{Term (*)}}, \tag{4.6}$$

where $\Lambda_T(k)$ is the quasi-likelihood test statistic for SCENARIO 1 defined in (3.2). It is proved in Theorem 3.5 that properly standardized statistic $\sup_{p < k \leq T} \Lambda_T(k)$ has asymptotically Gumbel distribution under Assumptions B and H_0 . In order we find the asymptotic distribution of the standardized maximum of (4.6) under H_0 , we will proceed as in Davis et al. (1995): Following their idea applied on univariate AR series, we should find process $\{T_k\}_{k \in \mathbb{Z}}$ such that it approximates Term (*) in (4.6) and the first differences $(T_k - T_{k-1})$

- are uncorrelated with the process $\{\mathbf{s}_k - \mathbf{s}_{k-1}\}_{k \in \mathbb{Z}}$,
- form a strong mixing sequence with the rate as stated in (B.3),
- have variance 2 - in order to be in line with the term $\frac{k-p}{2}$ appearing in the Taylor expansion.

Hence such process $\{T_k\}_{k \in \mathbb{Z}}$ should be relatively “simple” to satisfy the strong-mixing property but on the other hand it must not be too rigid to capture as much

4. DARLING-ERDÖS TEST FOR CHANGES IN VARIANCE

information as it is contained in Term (*) in (4.6). For univariate AR(p) processes, as shown in Davis et al. (1995), the second-order Taylor expansion of the quasi-likelihood test statistic is of the form (after discarding the asymptotically negligible terms)

$$\begin{aligned} & \frac{T-p}{\sigma^2} \cdot (\hat{\sigma}_T^2 - \sigma^2) - \frac{k-p}{\sigma^2} \cdot (\hat{\sigma}_k^2 - \sigma^2) - \frac{T-k}{\sigma^2} \cdot (\hat{\sigma}_k^2 - \sigma^2) + \\ & + \underbrace{\frac{k-p}{2\sigma^4} \cdot (\hat{\sigma}_k^2 - \sigma^2)^2}_{\text{Term (**)}}. \end{aligned} \quad (4.7)$$

Further, under conditions stated in Theorem 3.3 in Davis et al. (1995), these authors proposed the approximation of Term (**) in the Taylor expansion (4.7) by $T_k = \frac{1}{\sigma^2} \sum_{t=p+1}^k \varepsilon_t^2 - (k-p)$. They show that $\{T_k - T_{k-1}\}_{k \in \mathbb{Z}}$ forms a strong mixing sequence and

$$\frac{k-p}{2\sigma^4} \cdot (\hat{\sigma}_k^2 - \sigma^2)^2 - \frac{T_k^2}{2(k-p)} = \mathcal{O}\left(\frac{(\ln \ln k)^{\frac{3}{2}}}{k^{\frac{1}{2}}}\right) \text{ a.s., } k \rightarrow \infty.$$

It can therefore be seen that constant “2” that appears in the denominator of Term (**) in Taylor expression (4.7) is simultaneously the variance of the approximating term $(T_k - T_{k-1})$ provided that conditions in Theorem 3.3 of Davis et al. (1995) are met, especially that the first 4 moments of ε_t match to those of the standard normal random variable. Hence

$$\frac{T_k^2}{\text{var} T_k} - \frac{U_k^2}{2(k-p)} \xrightarrow[k \rightarrow \infty]{} 0 \text{ a.s.}$$

where $\{U_k\}_{k \in \mathbb{Z}}$ is an iid sequence coming from the FCLT. Now the Horváth’s extension of Darling-Erdős result can be used to obtain the desired Gumbel distribution, see Horváth (1993b).

This setup is difficult to perform in the multivariate case since term

$$\frac{k-p}{2} \text{tr}\{(\mathbf{\Omega}^{-1} \hat{\mathbf{\Omega}}_k - \mathbf{I}_n)^2\}$$

has too complicated structure to be approximated by a sum of a stationary strong mixing sequence. Even if we found such process $\{T_k\}_{k \in \mathbb{Z}}$, then $(T_k - T_{k-1})$ would not generally have variance equal to 2 even under normality conditions because of more complex structure of the variance in higher dimensions.

However we can consider modification of the form

$$\Lambda_T^* := \max_{p+1 \leq k \leq T} \{\Lambda_T^*(k)\}, \quad (4.8)$$

where

$$\Lambda_T^*(k) := \Lambda_T(k) + \frac{k-p}{\kappa - n^2} \left(\text{tr}\{\mathbf{\Omega}^{-1} \hat{\mathbf{\Omega}}_k - \mathbf{I}_n\} \right)^2, \quad (4.9)$$

$$\kappa := \mathbb{E} \left[\varepsilon_t^\top \mathbf{\Omega}^{-1} \varepsilon_t \right]^2. \quad (4.10)$$

In order to use this modification in practical application the estimation of $\mathbf{\Omega}$ and κ has to be sorted out. The properties of a suitable candidate $\hat{\mathbf{\Omega}}_T$ for estimation of $\mathbf{\Omega}$ is stated in Theorem 4.1.

4. DARLING-ERDÖS TEST FOR CHANGES IN VARIANCE

As regards κ , according to (B.2), κ does not depend on t hence we can write $\mathbb{E} \left[\boldsymbol{\varepsilon}_1^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_1 \right]^2$ instead of $\mathbb{E} \left[\boldsymbol{\varepsilon}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \right]^2$. For instance, if error term $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}} \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Omega})$ then $\kappa = 8$. Term κ is not known in practical applications but an estimate $\widehat{\kappa}_T$ such that $\widehat{\kappa}_T - \kappa = o\left((\ln \ln T)^{-1}\right)$ a.s., $T \rightarrow \infty$, can be used. We will show in Lemma 4.4, that estimate

$$\widehat{\kappa}_T = \frac{1}{T} \sum_{t=p+1}^T \left[\widehat{\boldsymbol{\varepsilon}}_t^\top \widehat{\boldsymbol{\Omega}}_T^{-1} \widehat{\boldsymbol{\varepsilon}}_t \right]^2 \quad (4.11)$$

fulfils this rate. To prove this, the appropriate speed of convergence in the strong law of large numbers is needed. Hence we will utilize FCLT B.10, even if it might seem too “brutal” for the reader to use FCLT for proving consistency. Another possibility is to use the law of the iterated logarithm. This option is briefly discussed below Assumptions B*. It can be seen from (4.11) that we will need to control higher moments of the error term up to order $8 + \delta$, $\delta > 0$, in order to utilize Theorem B.10. Hence the following modification of Assumptions B comes up:

ASSUMPTIONS B*:

Let condition (B.1) be valid and let error term process $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ be a sequence of centered random vectors. Assume further that

a*) assumption (B.2) a) holds,

b*) $\forall t_1 \leq t_2 \leq t_3, \forall (i_1, i_2, i_3): \mathbb{E}[\varepsilon_{i_1, t_1} \cdot \varepsilon_{i_2, t_2} \cdot \varepsilon_{i_3, t_3}] = 0$,

c*) assumption (B.2) c) holds and in addition, $\forall t_1 \leq \dots \leq t_6, \forall (i_1, \dots, i_6)$

$$\mathbb{E} \left[\varepsilon_{i_1, t_1} \cdot \dots \cdot \varepsilon_{i_6, t_6} \right] = \begin{cases} \mu_{i_1, \dots, i_6} & \text{if } t_1 = \dots = t_6, \\ \mu_{i_1, \dots, i_4} \sigma_{i_5 i_6} & \text{if } t_1 = \dots = t_4 < t_5 = t_6, \\ \sigma_{i_1 i_2} \sigma_{i_3 i_4} \sigma_{i_5 i_6} & \text{if } t_1 = t_2 < t_3 = t_4 < t_5 = t_6, \\ 0 & \text{otherwise,} \end{cases}$$

where μ_{i_1, \dots, i_6} is finite for all i_1, \dots, i_6 ;

further assume that $\forall t_1 \leq \dots \leq t_8, \forall (i_1, \dots, i_8)$

$$\mathbb{E} \left[\varepsilon_{i_1, t_1} \cdot \dots \cdot \varepsilon_{i_8, t_8} \right] = \begin{cases} \mu_{i_1, \dots, i_8} & \text{if } t_1 = \dots = t_8, \\ \mu_{i_1, \dots, i_6} \sigma_{i_7 i_8} & \text{if } t_1 = \dots = t_6 < t_7 = t_8, \\ \mu_{i_1, \dots, i_4} \mu_{i_5, \dots, i_8} & \text{if } t_1 = \dots = t_4 < \\ & < t_5 = \dots = t_8, \\ \mu_{i_1, \dots, i_4} \sigma_{i_5 i_6} \sigma_{i_7 i_8} & \text{if } t_1 = \dots = t_4 < \\ & < t_5 = t_6 < t_7 = t_8, \\ \sigma_{i_1 i_2} \sigma_{i_3 i_4} \sigma_{i_5 i_6} \sigma_{i_7 i_8} & \text{if } t_1 = t_2 < t_3 = t_4 < \\ & < t_5 = t_6 < t_7 = t_8, \\ 0 & \text{otherwise,} \end{cases}$$

and μ_{i_1, \dots, i_8} is finite for all i_1, \dots, i_8 ,

d*) $\sup_t \mathbb{E} \|\boldsymbol{\varepsilon}_t\|^{8+\delta} = \text{const.} < \infty$, for some $\delta > 0$.

Further suppose that process $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ is a strong mixing process with mixing rate $\rho_T = \mathcal{O}(T^{-(1+\epsilon)(1+8/\delta)})$ for some $\epsilon > 0$.

Instead of such complicated sets of conditions, it is sufficient to consider Assumptions B, strict stationarity of $\boldsymbol{\varepsilon}_t$, and b^*) and d^*). Law of the iterated logarithm for strictly stationary strong mixings, see Oodaira and Yoshihara (1971), Theorem 5, yields $\hat{\kappa}_T - \kappa = o((\ln \ln T)^{-1})$ a.s., $T \rightarrow \infty$. Condition b^*) stated above is necessary in order to get block-diagonal matrix in (4.44).

Test statistic $\Lambda_T^*(k)$ is simpler than $\Lambda_T'(k)$ since the square is taken outside of the trace operator. However, Λ_T^* still takes into account both diagonal and off-diagonal elements of the variance matrix and hence it can be used for the testing purposes. If H_0 is violated, the difference between the full-sample estimate $\hat{\boldsymbol{\Omega}}_T$ of the variance matrix $\boldsymbol{\Omega}$ and the estimate $\hat{\boldsymbol{\Omega}}_k$ based on the first $(k-p)$ observations is big which results in a large value of the test statistic. Therefore the large value of Λ_T^* in (4.8) signalizes violation of H_0 .

Let us rewrite the test statistic (4.9) for the univariate case ($n = 1$). Under assumption that $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, then $\kappa = \sigma^{-4} \cdot \mathbf{E}[\varepsilon_t^4] = 3$ is a kurtosis of standard normal random variable and hence $\kappa - n^2 = 2$. Hence rewriting (4.9) yields (4.7), i.e. the statistic considered by Davis et al. (1995). Hence it can be seen that our proposal Λ_T^* is a meaningful generalization into the more dimensions.

Next we will state the main convergence result for test statistic Λ_T^* under H_0 . The asymptotic distribution can be used for computation of the critical values and for the construction of the test. Under alternative, the simulation studies that have been conducted so far signalize the promising and desired result that the proposed test is consistent.

4.3 Main result

Theorem 4.1 *Let us assume that the VAR(p) model satisfies Assumptions B*. Let $\hat{\boldsymbol{\Omega}}_T$ be an estimate of $\boldsymbol{\Omega}$ such that*

$$\hat{\boldsymbol{\Omega}}_T - \boldsymbol{\Omega} = \mathcal{O}\left(T^{-\frac{1}{2}-\lambda}\right) \quad \text{a.s., for some } \lambda > 0, \quad \text{as } T \rightarrow \infty. \quad (4.12)$$

Then, under H_0 , it holds that

$$\mathbf{P}\left[\frac{\Lambda_T^* - b_T(d)}{a_T(d)} \leq x\right] \xrightarrow{T \rightarrow \infty} \exp\left\{-2e^{-\frac{x}{2}}\right\},$$

where $d = n(np + 1) + 1$ and

$$b_T(d) = \frac{\left(2 \ln \ln T + \frac{d}{2} \ln \ln \ln T - \ln \Gamma\left(\frac{d}{2}\right)\right)^2}{2 \ln \ln T},$$

$$a_T(d) = \sqrt{\frac{b_T(d)}{2 \ln \ln T}}.$$

The proof will be split into several parts later on. Notice that the size of constant d appearing in Theorem 4.1 is greater than the same constant in Theorem 3.5 by one. This reflects the fact that besides $\Lambda_T(k)$, we have additional

addend $\frac{k-p}{\kappa-n^2}(\text{tr}\{\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\Omega}}_k - \mathbf{I}_n\})^2$ in (4.9). To prove Theorem 4.1 a similar concept as in Davis et al. (1995) will be used. Some preliminary propositions needs to be proven first. Auxiliary propositions as well as Theorem 4.1 will be proven in forthcoming Section 4.4.

4.4 Proofs

Lemma 4.2 *If (elementwise)*

$$\mathbf{C}_T - \mathbf{C} = \mathcal{O}(T^{-\tau}) \quad \text{a.s.}, \quad T \rightarrow \infty,$$

holds for positive definite matrices \mathbf{C} , \mathbf{C}_T and constant $\tau > 0$, then also (elementwise)

$$\mathbf{C}_T^{-1} - \mathbf{C}^{-1} = \mathcal{O}(T^{-\tau}) \quad \text{a.s.}, \quad T \rightarrow \infty.$$

Proof: Proof is easy, however since it is often skipped in the literature we remind it here: Let $\tau > 0$. Then

$$\begin{aligned} \mathcal{O}(T^{-\tau}) &= \mathcal{O}(T^{-\tau}) + \mathcal{O}(T^{-2\tau}) \quad \text{a.s.} \\ \mathbf{I}_n - \mathbf{I}_n + \mathcal{O}(T^{-\tau}) &= (\mathbf{C} + \mathcal{O}(T^{-\tau})) \cdot \mathcal{O}(T^{-\tau}) \quad \text{a.s.} \\ \mathbf{I}_n - (\mathbf{C} + \mathcal{O}(T^{-\tau}))\mathbf{C}^{-1} &= \mathbf{C}_T \cdot \mathcal{O}(T^{-\tau}) \quad \text{a.s.} \\ \mathbf{C}_T\mathbf{C}_T^{-1} - \mathbf{C}_T\mathbf{C}^{-1} &= \mathbf{C}_T \cdot \mathcal{O}(T^{-\tau}) \quad \text{a.s.} \\ \mathbf{C}_T^{-1} - \mathbf{C}^{-1} &= \mathcal{O}(T^{-\tau}) \quad \text{a.s.}, \end{aligned}$$

where each line holds as $T \rightarrow \infty$. □

Lemma 4.3 *Under H_0 and Assumptions B it holds that*

$$\text{tr}\{\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\Omega}}_k\} = \frac{1}{k-p}\mathbf{e}_k^\top(\mathbf{I}_{k-p} \otimes \boldsymbol{\Omega}^{-1})\mathbf{e}_k + \mathcal{O}\left(\frac{\ln \ln k}{k}\right) \quad \text{a.s.}, \quad k \rightarrow \infty.$$

Proof: Due to (3.47) and the Law of the iterated logarithm applied to the sequence $\{\mathbf{u}_k\}_{k \in \mathbb{Z}}$, which was defined in the proof of Lemma 3.9, it holds that

$$\|\mathbf{s}_k\|^2 \leq \|\mathbf{s}_k - \mathbf{u}_k\|^2 + \|\mathbf{u}_k\|^2 = \mathcal{O}(k^{1-\lambda}) + \mathcal{O}(k \ln \ln k) \quad \text{a.s.}, \quad k \rightarrow \infty,$$

for some $\lambda > 0$. From that we obtain $\|\mathbf{s}_k\|^2 = \mathcal{O}(k \ln \ln k)$ a.s., $k \rightarrow \infty$.

Recall that, by (4.5), $Q_k = \mathbf{e}_k^\top(\mathbf{I}_{k-p} \otimes \boldsymbol{\Omega}^{-1})\mathbf{e}_k - \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k$. Applying Lemma 3.9, we get

$$Q_k = \mathbf{e}_k^\top(\mathbf{I}_{k-p} \otimes \boldsymbol{\Omega}^{-1})\mathbf{e}_k - \mathbf{u}_k^\top \frac{\boldsymbol{\Upsilon}^{-1}}{k} \mathbf{u}_k + o(1) \quad \text{a.s.}, \quad k \rightarrow \infty.$$

The Law of the iterated logarithm for iid centered random vectors \mathbf{u}_k yields $\mathbf{u}_k^\top (k\boldsymbol{\Upsilon})^{-1} \mathbf{u}_k = \mathcal{O}(\ln \ln k)$ a.s., $k \rightarrow \infty$. From that we obtain

$$Q_k = \mathbf{e}_k^\top(\mathbf{I}_{k-p} \otimes \boldsymbol{\Omega}^{-1})\mathbf{e}_k + \mathcal{O}(\ln \ln k) \quad \text{a.s.}, \quad k \rightarrow \infty. \quad (4.13)$$

Since $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \mathbf{M}_t \widehat{\boldsymbol{\beta}}_k$, then it holds

$$\begin{aligned} \text{tr}\left\{\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\Omega}}_k\right\} &= \text{tr}\left\{\boldsymbol{\Omega}^{-1}\frac{1}{k-p}\sum_{t=p+1}^k\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}_t^\top\right\} = \frac{1}{k-p}\sum_{t=p+1}^k\widehat{\boldsymbol{\varepsilon}}_t^\top\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\varepsilon}}_t = \\ &= \frac{Q_k}{k-p}\stackrel{(4.13)}{=} \frac{1}{k-p}\left(\mathbf{e}_k^\top(\mathbf{I}_{k-p}\otimes\boldsymbol{\Omega}^{-1})\mathbf{e}_k + \mathcal{O}(\ln\ln k)\right) \quad \text{a.s.}, \quad k \rightarrow \infty \\ &= \frac{1}{k-p}\mathbf{e}_k^\top(\mathbf{I}_{k-p}\otimes\boldsymbol{\Omega}^{-1})\mathbf{e}_k + \mathcal{O}\left(\frac{\ln\ln k}{k}\right) \quad \text{a.s.}, \quad k \rightarrow \infty. \end{aligned}$$

□

Lemma 4.4 *Let H_0 and Assumptions B^* be fulfilled. Then as $T \rightarrow \infty$,*

$$\widehat{\kappa}_T - \kappa = o\left((\ln\ln T)^{-1}\right) \quad \text{a.s.} \quad (4.14)$$

where $\widehat{\kappa}_T$ and κ is defined in (4.11) and in (4.10), respectively.

Proof: We will prove that $\widehat{\kappa}_T - \kappa = \mathcal{O}\left(\sqrt{\frac{\ln\ln T}{T}}\right)$ a.s. from which assertion (4.14) follows.

Due to Lemma A.2

$$\begin{aligned} &\frac{1}{T}\sum_{t=p+1}^T\left(\widehat{\boldsymbol{\varepsilon}}_t^\top\widehat{\boldsymbol{\Omega}}_T^{-1}\widehat{\boldsymbol{\varepsilon}}_t\right)^2 = \frac{1}{T}\sum_{t=p+1}^T\left(\left(\widehat{\boldsymbol{\varepsilon}}_t^\top\otimes\widehat{\boldsymbol{\varepsilon}}_t^\top\right)\text{vec}\left(\widehat{\boldsymbol{\Omega}}_T^{-1}\right)\right)^2 = \\ &= \frac{1}{T}\sum_{t=p+1}^T\text{tr}\left\{\left[\text{vec}\left(\widehat{\boldsymbol{\Omega}}_T^{-1}\right)\cdot\text{vec}\left(\widehat{\boldsymbol{\Omega}}_T^{-1}\right)^\top\right]\cdot\left[\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}_t^\top\otimes\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}_t^\top\right]\right\} = \\ &= \text{tr}\left\{\left[\text{vec}\left(\widehat{\boldsymbol{\Omega}}_T^{-1}\right)\cdot\text{vec}\left(\widehat{\boldsymbol{\Omega}}_T^{-1}\right)^\top\right]\cdot\frac{1}{T}\sum_{t=p+1}^T\left[\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}_t^\top\otimes\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}_t^\top\right]\right\}. \end{aligned} \quad (4.15)$$

Using $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \mathbf{M}_t\widehat{\boldsymbol{\beta}}_T = \boldsymbol{\varepsilon}_t - \mathbf{M}_t(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$, rule (5) in Lemma A.1, it holds that

$$\begin{aligned} &\frac{1}{T}\sum_{t=p+1}^T\left[\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}_t^\top\otimes\widehat{\boldsymbol{\varepsilon}}_t\widehat{\boldsymbol{\varepsilon}}_t^\top\right] = \\ &= \frac{1}{T}\sum_{t=p+1}^T\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\otimes\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right) - \end{aligned} \quad (4.16)$$

$$- \frac{1}{T}\sum_{t=p+1}^T\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\otimes\boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top\mathbf{M}_t^\top\right) - \quad (4.17)$$

$$- \frac{1}{T}\sum_{t=p+1}^T\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\otimes\mathbf{M}_t(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})\boldsymbol{\varepsilon}_t^\top\right) + \quad (4.18)$$

$$+ \frac{1}{T}\sum_{t=p+1}^T\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\otimes\mathbf{M}_t(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top\mathbf{M}_t^\top\right) - \quad (4.19)$$

$$- \frac{1}{T}\sum_{t=p+1}^T\left(\boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top\mathbf{M}_t^\top\otimes\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right) + \quad (4.20)$$

$$+ \frac{1}{T}\sum_{t=p+1}^T\left(\boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top\mathbf{M}_t^\top\otimes\boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top\mathbf{M}_t^\top\right) + \quad (4.21)$$

$$+ \frac{1}{T} \sum_{t=p+1}^T \left(\boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \otimes \mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \boldsymbol{\varepsilon}_t^\top \right) - \quad (4.22)$$

$$- \frac{1}{T} \sum_{t=p+1}^T \left(\boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \otimes \mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) - \quad (4.23)$$

$$- \frac{1}{T} \sum_{t=p+1}^T \left(\mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \boldsymbol{\varepsilon}_t^\top \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \right) + \quad (4.24)$$

$$+ \frac{1}{T} \sum_{t=p+1}^T \left(\mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \boldsymbol{\varepsilon}_t^\top \otimes \boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) + \quad (4.25)$$

$$+ \frac{1}{T} \sum_{t=p+1}^T \left(\mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \boldsymbol{\varepsilon}_t^\top \otimes \mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \boldsymbol{\varepsilon}_t^\top \right) - \quad (4.26)$$

$$- \frac{1}{T} \sum_{t=p+1}^T \left(\mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \boldsymbol{\varepsilon}_t^\top \otimes \mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) + \quad (4.27)$$

$$+ \frac{1}{T} \sum_{t=p+1}^T \left(\mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \right) - \quad (4.28)$$

$$- \frac{1}{T} \sum_{t=p+1}^T \left(\mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \otimes \boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) - \quad (4.29)$$

$$- \frac{1}{T} \sum_{t=p+1}^T \left(\mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \otimes \mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \boldsymbol{\varepsilon}_t^\top \right) + \quad (4.30)$$

$$+ \frac{1}{T} \sum_{t=p+1}^T \left(\mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \otimes \mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) \quad (4.31)$$

We will investigate the asymptotics of each matrices element-by-element and hence it does not matter whether we investigate its elements stacked in the vector form or in the matrix composition. It is due to the fact that all matrices have fixed dimensions even when $T \rightarrow \infty$. Therefore, without loss of generality, we can plug in vec operator inside all sums and in cases where it is convenient. This transformation does not influence the results.

Let us focus on particular summands: We can apply FCLT B.10 on the vec of expression (4.16) to get

$$\frac{1}{T} \sum_{t=p+1}^T \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \right) = \mathbf{E} \left[\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^\top \otimes \boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^\top \right] + \mathcal{O} \left(\sqrt{\frac{\ln \ln T}{T}} \right) \quad \text{a.s.}, \quad T \rightarrow \infty.$$

Applying vec on expression (4.17), we get

$$\begin{aligned} & \frac{1}{T} \sum_{t=p+1}^T \text{vec} \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \otimes \boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) = \\ & = \frac{1}{T} \sum_{t=p+1}^T \text{vec} \left(\left(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t \right) \left(\mathbf{1} \boldsymbol{\varepsilon}_t^\top \otimes (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) \right) = \\ & = \frac{1}{T} \sum_{t=p+1}^T \text{vec} \left(\left(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t \right) \left(\mathbf{1} \otimes (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \right) \left(\boldsymbol{\varepsilon}_t^\top \otimes \mathbf{M}_t^\top \right) \right) = \\ & = \frac{1}{T} \sum_{t=p+1}^T \left(\left(\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t \right) \otimes \left(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t \right) \right) \cdot \text{vec} \left((\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \right). \end{aligned}$$

4. DARLING-ERDÖS TEST FOR CHANGES IN VARIANCE

It is not difficult to show that $\mathbf{E} \left[\text{vec} \left((\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t) \otimes (\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t) \right) \right] = \mathbf{0} \in \mathbb{R}^{n^5(np+1)}$, for all t . Using Theorem B.10 on $\text{vec} \left((\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t) \otimes (\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t) \right)$ and (3.20) we obtain that

$$\frac{1}{T} \sum_{t=p+1}^T \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \otimes \boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) = \mathcal{O} \left(\frac{\ln \ln T}{T} \right) \quad \text{a.s.}, \quad T \rightarrow \infty.$$

Terms (4.18), (4.20), (4.24) can be treated in the same way to get rate $\mathcal{O} \left(\frac{\ln \ln T}{T} \right)$ a.s., as $T \rightarrow \infty$.

Analogously, applying vec operator on term (4.19) we get after some algebra

$$\frac{1}{T} \sum_{t=p+1}^T \left((\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t) \otimes (\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t) \right) \cdot \text{vec} \left((\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \right).$$

Since $\mathbf{E} \left[\text{vec} \left((\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t) \otimes (\boldsymbol{\varepsilon}_t \otimes \mathbf{M}_t) \right) \right] = \mathbf{E} \left[\text{vec} \left((\boldsymbol{\varepsilon}_1 \otimes \mathbf{M}_1) \otimes (\boldsymbol{\varepsilon}_1 \otimes \mathbf{M}_1) \right) \right] < \infty$, for all t and using (3.20) we get by Theorem B.10 that as $T \rightarrow \infty$,

$$\begin{aligned} & \frac{1}{T} \sum_{t=p+1}^T \text{vec} \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \otimes \mathbf{M}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) = \\ & = \left[\mathbf{E} \left[(\boldsymbol{\varepsilon}_1 \otimes \mathbf{M}_1) \otimes (\boldsymbol{\varepsilon}_1 \otimes \mathbf{M}_1) \right] + \mathcal{O} \left(\sqrt{\frac{\ln \ln T}{T}} \right) \right] \cdot \mathcal{O} \left(\frac{\ln \ln T}{T} \right) \quad \text{a.s.} \\ & = \mathcal{O} \left(\frac{\ln \ln T}{T} \right) \quad \text{a.s.} \end{aligned}$$

Term (4.28) can be treated in the same way and achieve rate $\mathcal{O} \left(\frac{\ln \ln T}{T} \right)$ a.s., as $T \rightarrow \infty$.

Applying vec operator on term (4.21), the algebraic computations lead to

$$\begin{aligned} & \frac{1}{T} \sum_{t=p+1}^T \text{vec} \left(\boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \otimes \boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) = \\ & = \frac{1}{T} \sum_{t=p+1}^T \left(\mathbf{M}_t \otimes \mathbf{M}_t \otimes \boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t \right) \cdot \text{vec} \left((\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \otimes (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \right). \end{aligned}$$

Using FCLT B.10 on $\text{vec}(\mathbf{M}_t \otimes \mathbf{M}_t \otimes \boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t)$ and (3.20) we get that

$$\begin{aligned} & \frac{1}{T} \sum_{t=p+1}^T \left(\boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \otimes \boldsymbol{\varepsilon}_t (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})^\top \mathbf{M}_t^\top \right) = \\ & = \mathcal{O} \left(\frac{\ln \ln T}{T} \right) \quad \text{a.s.}, \quad T \rightarrow \infty \end{aligned}$$

and the same rate is achieved also for terms (4.22), (4.25) and (4.26).

Applying similar setup on the rest of the terms we can conclude that terms (4.23), (4.27), (4.29) and (4.30) are $\mathcal{O} \left(\left(\frac{\ln \ln T}{T} \right)^{\frac{3}{2}} \right)$ a.s., $T \rightarrow \infty$ and term (4.31) is $\mathcal{O} \left(\left(\frac{\ln \ln T}{T} \right)^2 \right)$ a.s., $T \rightarrow \infty$.

Using (4.15), (3.21) and above results

$$\begin{aligned}
 \widehat{\kappa}_T - \kappa &= \text{tr} \left\{ \left[\text{vec}(\widehat{\Omega}_T^{-1}) \cdot \text{vec}(\widehat{\Omega}_T^{-1})^\top \right] \cdot \left[\frac{1}{T} \sum_{t=p+1}^T (\widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t^\top \otimes \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t^\top) \right] \right\} = \\
 &= \text{tr} \left\{ \left[\text{vec}(\Omega^{-1}) \cdot \text{vec}(\Omega^{-1})^\top + \mathcal{O}\left(\frac{\ln \ln T}{T}\right) \right] \cdot \right. \\
 &\quad \left. \cdot \left[\mathbb{E} \left[\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^\top \otimes \boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^\top \right] + \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \right] \right\} \quad \text{a.s.}, \quad T \rightarrow \infty \\
 &= \mathbb{E} \text{tr} \left\{ \left[\text{vec}(\Omega^{-1}) \cdot \text{vec}(\Omega^{-1})^\top \cdot \left[\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^\top \otimes \boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^\top \right] \right] \right\} + \\
 &\quad + \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \quad \text{a.s.}, \quad T \rightarrow \infty \\
 &= \mathbb{E} \left[\boldsymbol{\varepsilon}_1^\top \Omega^{-1} \boldsymbol{\varepsilon}_1 \right]^2 + \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \quad \text{a.s.}, \quad T \rightarrow \infty,
 \end{aligned}$$

which completes the proof. \square

Lemma 4.5 *Let H_0 and Assumptions B^* be fulfilled and let*

$$T_k := \text{tr} \left\{ \Omega^{-1} \sum_{t=p+1}^k \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - (k-p) \mathbf{I}_n \right\}. \quad (4.32)$$

Let us suppose that $\widehat{\Omega}_T$ fulfils (4.12). Then, as $T \rightarrow \infty$

$$\frac{k-p}{\widehat{\kappa}_T - n^2} \left(\text{tr} \{ \widehat{\Omega}_T^{-1} \widehat{\Omega}_k - \mathbf{I}_n \} \right)^2 - \frac{T_k^2}{(\kappa - n^2)(k-p)} = o(1) \quad \text{a.s.} \quad (4.33)$$

Proof: The difference in statement (4.33) can be decomposed as follows:

$$\frac{k-p}{\widehat{\kappa}_T - n^2} \left(\text{tr} \{ \widehat{\Omega}_T^{-1} \widehat{\Omega}_k - \mathbf{I}_n \} \right)^2 - \frac{k-p}{\kappa - n^2} \left(\text{tr} \{ \widehat{\Omega}_T^{-1} \widehat{\Omega}_k - \mathbf{I}_n \} \right)^2 + \quad (4.34)$$

$$+ \frac{k-p}{\kappa - n^2} \left(\text{tr} \{ \widehat{\Omega}_T^{-1} \widehat{\Omega}_k - \mathbf{I}_n \} \right)^2 - \frac{k-p}{\kappa - n^2} \left(\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k - \mathbf{I}_n \} \right)^2 + \quad (4.35)$$

$$+ \frac{k-p}{\kappa - n^2} \left(\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k - \mathbf{I}_n \} \right)^2 - \frac{T_k^2}{(\kappa - n^2)(k-p)}. \quad (4.36)$$

Term T_k can be expressed as

$$T_k = \sum_{t=p+1}^k (\boldsymbol{\varepsilon}_t^\top \Omega^{-1} \boldsymbol{\varepsilon}_t - n) = (k-p) \cdot \left(\frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k - n \right). \quad (4.37)$$

Step 1:

First, we will prove that terms (4.36) satisfy

$$\frac{k-p}{\kappa - n^2} \left(\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k - \mathbf{I}_n \} \right)^2 - \frac{T_k^2}{(\kappa - n^2)(k-p)} = \mathcal{O}\left(\frac{(\ln \ln k)^{\frac{3}{2}}}{k^{\frac{1}{2}}}\right) \quad \text{a.s.}, \quad (4.38)$$

4. DARLING-ERDÖS TEST FOR CHANGES IN VARIANCE

$k \rightarrow \infty$. Proof of the latter statement is based on the following set of computations:

$$\begin{aligned}
& \frac{k-p}{\kappa-n^2} \left(\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k - \mathbf{I}_n \} \right)^2 - \frac{T_k^2}{(\kappa-n^2)(k-p)} = \\
&= \frac{k-p}{\kappa-n^2} \left(\left[\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k - \mathbf{I}_n \} \right]^2 - \left[\frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k - n \right]^2 \right) = \\
&= \frac{k-p}{\kappa-n^2} \left(\left[\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k \} - n + \frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k - n \right] \cdot \right. \\
&\quad \left. \cdot \left[\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k \} - n - \frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k + n \right] \right) = \\
&= \frac{k-p}{\kappa-n^2} \left(\left[\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k \} + \frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k - 2n \right] \cdot \right. \\
&\quad \left. \cdot \left[\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k \} - \frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k \right] \right) =: D.
\end{aligned}$$

Applying the approximation to $\text{tr} \{ \Omega^{-1} \widehat{\Omega}_k \}$ from Lemma 4.3 we obtain that

$$\begin{aligned}
D &= \frac{k-p}{\kappa-n^2} \left(\left[\frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k + \mathcal{O}\left(\frac{\ln \ln k}{k}\right) + \right. \right. \\
&\quad \left. \left. + \frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k - 2n \right] \cdot \left[\frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k + \right. \right. \\
&\quad \left. \left. + \mathcal{O}\left(\frac{\ln \ln k}{k}\right) - \frac{1}{k-p} \mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k \right] \right) \quad \text{a.s. } k \rightarrow \infty \\
&= 2 \cdot \frac{k-p}{\kappa-n^2} \left(\frac{1}{k-p} \left[\mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k - n(k-p) \right] + \mathcal{O}\left(\frac{\ln \ln k}{k}\right) \right) \cdot \\
&\quad \cdot \mathcal{O}\left(\frac{\ln \ln k}{k}\right) \quad \text{a.s. } k \rightarrow \infty \\
&= 2 \cdot \frac{k-p}{\kappa-n^2} \left(\frac{1}{k-p} \mathcal{O}(\sqrt{k \ln \ln k}) + \mathcal{O}\left(\frac{\ln \ln k}{k}\right) \right) \cdot \mathcal{O}\left(\frac{\ln \ln k}{k}\right) \quad \text{a.s. } k \rightarrow \infty \\
&= \mathcal{O}\left(\frac{(\ln \ln k)^{\frac{3}{2}}}{k^{\frac{1}{2}}}\right) \quad \text{a.s. } k \rightarrow \infty.
\end{aligned}$$

The last-but-one equality follows from the Law of the iterated logarithm:

$$\begin{aligned}
\mathbf{e}_k^\top (\mathbf{I}_{k-p} \otimes \Omega^{-1}) \mathbf{e}_k - n(k-p) &= \sum_{t=p+1}^k \left(\boldsymbol{\varepsilon}_t^\top \Omega^{-1} \boldsymbol{\varepsilon}_t - n \right) = \\
&= \mathcal{O}(\sqrt{k \ln \ln k}) \quad \text{a.s., } k \rightarrow \infty.
\end{aligned}$$

Step 2:

As regards terms in (4.35), the difference is equal to

$$\frac{k-p}{\kappa-n^2} \cdot \text{tr} \left\{ \left(\widehat{\Omega}_T^{-1} + \Omega^{-1} \right) \widehat{\Omega}_k - 2\mathbf{I}_n \right\} \cdot \text{tr} \left\{ \left(\widehat{\Omega}_T^{-1} - \Omega^{-1} \right) \widehat{\Omega}_k \right\}.$$

Now, due to Lemma 4.2, Lemma 3.8 and Assumption (4.12)

$$\begin{aligned}
\left(\widehat{\Omega}_T^{-1} - \Omega^{-1} \right) \widehat{\Omega}_k &= \left(\widehat{\Omega}_T^{-1} - \Omega^{-1} \right) \left(\widehat{\Omega}_k - \Omega \right) + \left(\widehat{\Omega}_T^{-1} - \Omega^{-1} \right) \Omega = \\
&= \mathcal{O}\left(\frac{\sqrt{\ln \ln k}}{k^{1+\lambda}}\right) + \mathcal{O}\left(k^{-\frac{1}{2}-\lambda}\right) \quad \text{a.s.} = \\
&= \mathcal{O}\left(k^{-\frac{1}{2}-\lambda}\right) \quad \text{a.s., for some } \lambda > 0
\end{aligned}$$

and

$$\begin{aligned} (\widehat{\Omega}_T^{-1} + \Omega^{-1})\widehat{\Omega}_k - 2\mathbf{I}_n &= (\widehat{\Omega}_T^{-1} - \Omega^{-1})(\widehat{\Omega}_k - \Omega) + 2\Omega^{-1}(\widehat{\Omega}_k - \Omega) + \\ &\quad + (\widehat{\Omega}_T^{-1} - \Omega^{-1})\Omega = \mathcal{O}\left(\sqrt{\frac{\ln \ln k}{k}}\right) \quad \text{a.s.} \end{aligned}$$

Hence,

$$\text{tr}\left\{(\widehat{\Omega}_T^{-1} + \Omega^{-1})\widehat{\Omega}_k - 2\mathbf{I}_n\right\} \cdot \text{tr}\left\{(\widehat{\Omega}_T^{-1} - \Omega^{-1})\widehat{\Omega}_k\right\} = \mathcal{O}\left(\frac{\sqrt{\ln \ln k}}{k^{1+\lambda}}\right) \quad \text{a.s.}$$

for some $\lambda > 0$. We have just shown that, for (4.35) it holds

$$\begin{aligned} &\frac{k-p}{\kappa-n^2} \left(\text{tr}\{\widehat{\Omega}_T^{-1}\widehat{\Omega}_k - \mathbf{I}_n\}\right)^2 - \frac{k-p}{\kappa-n^2} \left(\text{tr}\{\Omega^{-1}\widehat{\Omega}_k - \mathbf{I}_n\}\right)^2 = \\ &= \mathcal{O}\left(\frac{\sqrt{\ln \ln k}}{k^\lambda}\right), \quad \text{a.s., } k \rightarrow \infty, \end{aligned} \quad (4.39)$$

for some $\lambda > 0$.

Step 3:

Let us conclude the proof with the terms in (4.34) which is equal to

$$\left[\frac{k-p}{\widehat{\kappa}-n^2} - \frac{k-p}{\kappa-n^2}\right] \cdot \left(\text{tr}\{\widehat{\Omega}_T^{-1}\widehat{\Omega}_k - \mathbf{I}_n\}\right)^2.$$

We investigate each of the brackets separately:

$$\frac{k-p}{\widehat{\kappa}-n^2} - \frac{k-p}{\kappa-n^2} = \frac{(k-p)(\kappa-\widehat{\kappa})}{(\widehat{\kappa}-\kappa)(\kappa-n^2) + (\kappa-n^2)^2} = o\left(\frac{k}{\ln \ln k}\right) \quad \text{a.s., } k \rightarrow \infty,$$

due to Lemma 4.4 and the fact that $(\kappa-n^2)^2 > 0$. Lemma 4.2, Lemma 3.8 and Assumption (4.12) yield

$$\begin{aligned} \widehat{\Omega}_T^{-1}\widehat{\Omega}_k - \mathbf{I}_n &= (\widehat{\Omega}_T^{-1} - \Omega^{-1})(\widehat{\Omega}_k - \Omega) + (\widehat{\Omega}_T^{-1} - \Omega^{-1})\Omega + \Omega^{-1}(\widehat{\Omega}_k - \Omega) = \\ &= \mathcal{O}\left(\sqrt{\frac{\ln \ln k}{k}}\right) \quad \text{a.s. } k \rightarrow \infty, \end{aligned}$$

and hence

$$\left(\text{tr}\{\widehat{\Omega}_T^{-1}\widehat{\Omega}_k - \mathbf{I}_n\}\right)^2 = \mathcal{O}\left(\frac{\ln \ln k}{k}\right) \quad \text{a.s. } k \rightarrow \infty,$$

which implies that (4.34) is $o(1)$ a.s. This result together with (4.38) and (4.39) yields the assertion of the lemma. \square

Proof of Theorem 4.1: According to Theorem 3.5, the normalized version of Λ_T

4. DARLING-ERDÖS TEST FOR CHANGES IN VARIANCE

has asymptotically the Gumbel distribution, as $T \rightarrow \infty$. We have to show that

$$\lim_{\epsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P} \left[\left| \max_{p < k \leq T\epsilon} \Lambda_T^*(k) - \max_{p < k \leq T\epsilon} \left\{ \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k + \frac{T_k^2}{(k-p)(\kappa - n^2)} \right\} \right| > \delta \right] = 0, \quad (4.40)$$

$$\lim_{\epsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P} \left[\left| \max_{(1-\epsilon)T \leq k < T} \Lambda_T^*(k) - \max_{(1-\epsilon)T \leq k < T} \left\{ \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k + \frac{T_k^2}{(k-p)(\kappa - n^2)} \right\} \right| > \delta \right] = 0, \quad (4.41)$$

$$\lim_{\epsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P} \left[\frac{\max_{p < k \leq T\epsilon} \left\{ \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k + \frac{T_k^2}{(k-p)(\kappa - n^2)} \right\} - b_T(d)}{a_T(d)} \leq x \right] - \exp \left\{ -e^{-\frac{x}{2}} \right\} = 0, \quad (4.42)$$

$$\lim_{\epsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P} \left[\frac{\max_{(1-\epsilon)T \leq k < T} \left\{ \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{s}}_k + \frac{T_k^2}{(k-p)(\kappa - n^2)} \right\} - b_T(d)}{a_T(d)} \leq x \right] - \exp \left\{ -e^{-\frac{x}{2}} \right\} = 0, \quad (4.43)$$

where T_k is defined in (4.32) or (4.37). To establish (4.40) observe that

$$\begin{aligned} & \left| \max_{p < k \leq T\epsilon} \Lambda_T^*(k) - \max_{p < k \leq T\epsilon} \left\{ \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k + \frac{T_k^2}{(k-p)(\kappa - n^2)} \right\} \right| \leq \\ & \leq \max_{p < k \leq T\epsilon} \left| \Lambda_T(k) - \mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k \right| + \\ & + \max_{p < k \leq T\epsilon} \left| \frac{k-p}{\hat{\kappa}_T - n^2} \left(\text{tr} \{ \hat{\boldsymbol{\Omega}}_T^{-1} \hat{\boldsymbol{\Omega}}_k - \mathbf{I}_n \} \right)^2 - \frac{T_k^2}{(k-p)(\kappa - n^2)} \right|. \end{aligned}$$

Result (4.40) now follows from the proof of Lemma 3.10 and Lemma 4.5. Proof of (4.41) can be handled similarly. We establish (4.42). Define vector

$$\begin{aligned} \boldsymbol{\xi}_k & := \begin{pmatrix} \mathbf{s}_k - \mathbf{s}_{k-1} \\ T_k - T_{k-1} \end{pmatrix} = \begin{pmatrix} \mathbf{M}_k^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_k \\ \boldsymbol{\varepsilon}_k^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_k - n \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1(\mathbf{y}_k, \dots, \mathbf{y}_{k-p}) \\ f_2(\mathbf{y}_k, \dots, \mathbf{y}_{k-p}) \end{pmatrix} = \\ & = \mathbf{f}(\mathbf{y}_k, \dots, \mathbf{y}_{k-p}) \end{aligned}$$

which is a measurable function of finite strong mixing terms $\mathbf{y}_k, \dots, \mathbf{y}_{k-p}$ and according to Theorem B.9 it is a strong mixing sequence with the same rate as \mathbf{y}_t . The variance matrix of $\boldsymbol{\xi}$ is

$$\boldsymbol{\Gamma} := \begin{pmatrix} \boldsymbol{\Upsilon} & \mathbf{0} \\ \mathbf{0}^\top & \kappa - n^2 \end{pmatrix}, \quad \text{where } \boldsymbol{\Upsilon} = \mathbb{E} \left[\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \mathbf{M}_t \right]. \quad (4.44)$$

First differences $\mathbf{s}_k - \mathbf{s}_{k-1}$ and $T_k - T_{k-1}$ are uncorrelated thanks to assumption $\mathbb{E} [\varepsilon_{i,t} \varepsilon_{j,t} \varepsilon_{k,t}] = 0$, $\forall (i, j, k), \forall t$. According to Theorem B.10 there exist a sequence of iid centered Gaussian random elements $(\mathbf{z}_{k,1}^\top, z_{k,2}^\top)^\top$ with covariance matrix $\boldsymbol{\Gamma}$ such that as $k \rightarrow \infty$

$$\begin{pmatrix} \mathbf{s}_k \\ T_k \end{pmatrix} - \begin{pmatrix} \mathbf{u}_{k,1} \\ u_{k,2} \end{pmatrix} = \mathcal{O}(k^{\frac{1}{2}-\lambda}) \quad \text{a.s. for some } \lambda > 0 \quad (4.45)$$

where $\mathbf{u}_{k,1} = \sum_{t=1}^k \mathbf{z}_{k,1}$ and $u_{k,2} = \sum_{t=1}^k z_{k,2}$. Due to Lemma 3.9

$$\mathbf{s}_k^\top \mathbf{P}_k \mathbf{s}_k - \mathbf{u}_{k,1}^\top \frac{\mathbf{\Upsilon}^{-1}}{k} \mathbf{u}_{k,1} \xrightarrow[k \rightarrow \infty]{} 0. \quad \text{a.s.}$$

Now, due to the Law of the iterated logarithm $u_{k,2} = \mathcal{O}(\sqrt{k \ln \ln k})$ a.s., $k \rightarrow \infty$, and due to (4.45) we have

$$\begin{aligned} & \frac{T_k^2}{(k-p)(\kappa-n^2)} - \frac{u_{k,2}^2}{(k-p)(\kappa-n^2)} = \frac{(T_k - u_{k,2})^2 + 2u_{k,2}(T_k - u_{k,2})}{(k-p)(\kappa-n^2)} = \\ & = \frac{\mathcal{O}(k^{1-\lambda}) + \sqrt{k \ln \ln k} \cdot \mathcal{O}(k^{\frac{1}{2}-\lambda})}{(k-p)(\kappa-n^2)} \quad \text{a.s. for some } \lambda > 0, k \rightarrow \infty \\ & = \mathcal{O}\left(\frac{\sqrt{\ln \ln k}}{k^\lambda}\right) \quad \text{a.s. for some } \lambda > 0, k \rightarrow \infty \\ & = o(1) \quad \text{a.s., } k \rightarrow \infty, \end{aligned}$$

and hence

$$\frac{T_k^2}{(k-p)(\kappa-n^2)} - \frac{u_{k,2}^2}{(k-p)(\kappa-n^2)} = o(1) \quad \text{a.s., } k \rightarrow \infty.$$

The rest of the proof follows from the strong approximation theorem applied to the iid random variables $u_{k,2}$, see Horváth (1993b), Lemma 2.2. The proof of (4.43) follows from the similar considerations as the just finished proof of (4.42). \square

4.5 Simulation study

A simulation study illustrates the performance of the proposed testing scheme. It is worth noting here that the convergence to the Gumbel distribution is rather slow which was confirmed also for instance in Davison (2003) or Horváth (1993b) in much simpler cases than multivariate autoregressions. Satisfactory results are achieved for large sample sizes. Hence we will not simulate here the situation under alternative hypothesis due to the fact that neither results under H_0 are satisfactory. To get more reliable critical values one might use bootstrapping techniques, see Hušková et al. (2008) for instance.

The same testing scheme will be kept as in the previous chapter: We will consider processes P_1 and P_2 defined on Page 39 with errors [E1] as defined in Section 3.5. As regards [E2] error term, the degrees of freedom ν in t -distribution will be increased to 10 in order to fulfil the Assumptions B*.

The test statistic is more complicated than in case of no changes in variance of errors, hence it is more time-consuming to get the results. The calculations were done in software R as earlier described.

We will investigate the distribution of

$$\frac{\Lambda_T^* - b_T(n(np+1)+1)}{a_T(n(np+1)+1)}, \quad n=2, p=1. \quad (4.46)$$

4. DARLING-ERDÖS TEST FOR CHANGES IN VARIANCE

Figure 4.2 gives the comparison of the distribution function of (4.46) based on 1 000 simulations for different lengths $T = 100, 1\,000, 10\,000$ and asymptotic Gumbel distribution as given in Theorem 4.1. Good asymptotic results are achieved only for large sample sizes.

According to Figure 4.2 the asymptotic distribution tends to be smaller than the replicates of the test statistic (4.46). Let us focus on the properties of the estimators appearing in Λ_T^* . The convergence of standardized maximum of $\Lambda_T(k)$ to Gumbel distribution is slow, see Dvořák and Prášková (2013). The other terms that can negatively influence the speed of convergence of Λ_T^* are $\widehat{\kappa}_T$ and $\widehat{\Omega}_T^{-1}$. Simulations do not give evidence on poorer convergence of $\widehat{\Omega}_T^{-1}$ to Ω^{-1} .

Let us concentrate on term $\widehat{\kappa}_T$. In Figure 4.1, results of 100 simulations were plotted for the estimators of κ for different $T = 100, 1\,000, 10\,000$. In the top 2 plots the error term is assumed to have Gaussian distribution and in the bottom 2 figures, more heavy-tailed t_{10} distribution with the parameter $\frac{\nu-2}{\nu}\Omega$ is chosen (i.e. the resulting variance is the same as for the Gaussian case). It can be seen that in both cases the estimators correctly oscillate around the true value $\kappa = 8$ (top panels), $\kappa = 8 \cdot \frac{\nu-2}{\nu-4} = \frac{32}{3} \doteq 10.67$ (bottom panels), respectively, see calculation in Kotz and Nadarajah (2004) for higher moments of t -distribution. There is a somewhat higher standard deviation in the bottom figures and also for smaller sample sizes in both panels. In Table 4.1 there are results based on 1 000 simulations.

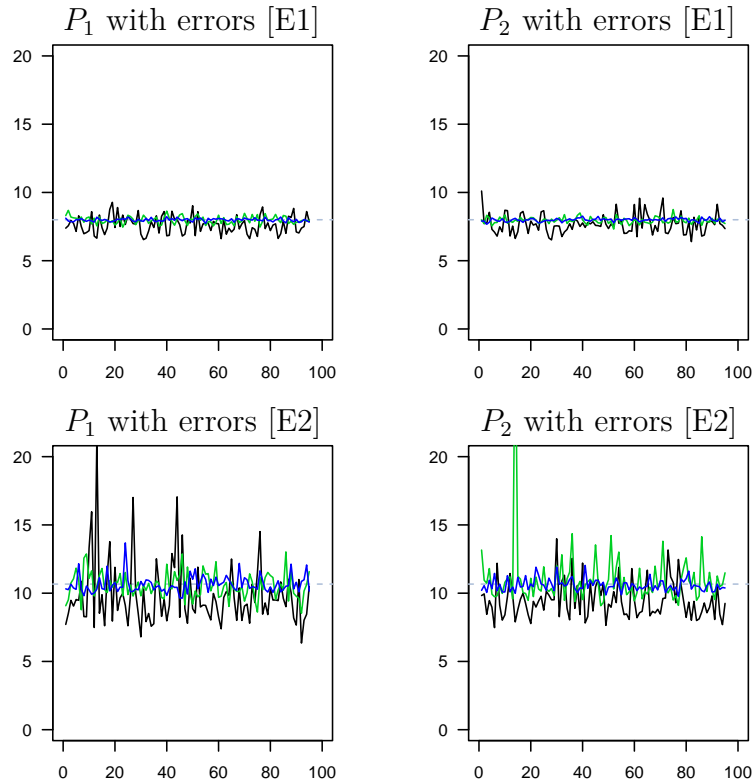


Figure 4.1: Estimator $\widehat{\kappa}_T$ for each of 100 repetitions. Gaussian (top panels) and $t_{10}(\mathbf{0}, \frac{\nu-2}{\nu}\Omega)$ distribution (bottom panels) considered for the error term process. Black, green and blue color stands for $T = 100, T = 1\,000$ and for $T = 10\,000$, respectively.

Table 4.1: Simulated values of $\hat{\kappa}_T$ for different T . Errors [E1], [E2] considered.

T	[E1]				[E2]			
	$\hat{\kappa}_T$	P_1 std($\hat{\kappa}_T$)	$\hat{\kappa}_T$	P_2 std($\hat{\kappa}_T$)	$\hat{\kappa}_T$	P_1 std($\hat{\kappa}_T$)	$\hat{\kappa}_T$	P_2 std($\hat{\kappa}_T$)
100	7.73	0.728	7.75	0.746	9.65	2.027	9.62	2.054
1 000	7.97	0.253	7.97	0.243	10.50	1.086	10.60	1.692
10 000	8.00	0.114	7.99	0.111	10.63	0.722	10.63	0.507

It can be seen from Table 4.1 that κ is underestimated on average in smaller samples. This might result in slightly higher values of Λ_T^* since $\hat{\kappa}_T$ appears as a positive term in the denominator of Λ_T^* . This can give a partial explanation why the distribution of (4.46) is shifted to the right from the Gumbel distribution in Figure 4.2.

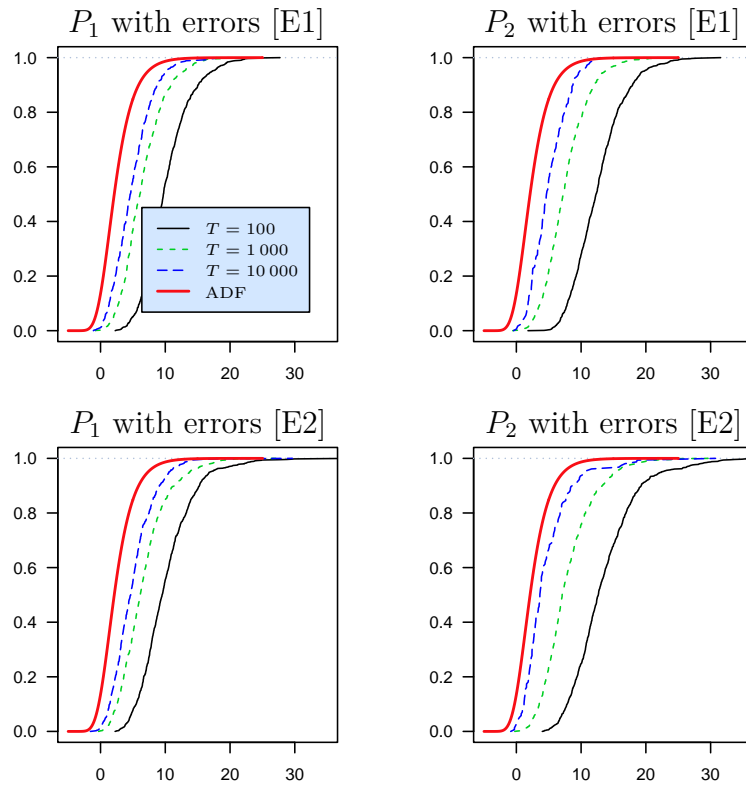


Figure 4.2: Empirical and Asymptotic distribution functions.

We checked that the diagonalization of $\mathbf{\Omega}$ does not help in reaching better results.

4.6 Chapter Summary

We conclude that this chapter is interesting merely from the theoretical point of view. We have shown that the natural generalization of standardized quasi-likelihood ratio test statistic from the univariate case does not follow Gumbel

distribution under H_0 and have given the reason which shows up under Taylor expansion of the test statistic. A modification has been presented in order to get some satisfactory results under H_0 , however, a better estimate $\hat{\Omega}_T$ is needed in order the desired convergence could have been proven. Also to get a consistent estimate of κ one needs higher moment conditions under weak dependence than it is usual in the change point literature.

As regards the short simulation study, we conclude that the speed of convergence is rather very slow. Besides the dimensionality, the quality of convergence is sensitive especially on the persistence of VAR(p) model and the quality of the estimator of κ which converge to the true value slowly as well. The proposed testing scheme can be applied on stationary data of very large sample sizes. Unfortunately, author does not have enough time for performing bootstrapping comparisons which might be useful in this case and this gives a possibility for future investigations.

5. Score test

5.1 Introduction

In this section we will discuss a *score test* which is as well as likelihood-ratio-type tests derived from the pseudo-likelihood function. The idea of the score test statistic goes back to Fisher (1925) where the efficient statistic based on the first derivatives of the likelihood is derived. Also the score test presented here is based on the first derivatives of the likelihood function. Such type of test can be used both in the sequential monitoring of possible change-points, see for instance Gombay and Serban (2009), and also for the retrospective testing, i.e. Gombay (2008). The contents of Chapter 5 is based on the latter article however the theory will be extended to multivariate stationary autoregressions. Compared to Gombay (2008), we will present the proofs in more detail, since it is very difficult to follow some logical steps in the latter article. Main advantages of our approach over the latter paper are twofold: First, the ability to test multivariate autoregression, and, second, to use the test also for non-Gaussian distributions.

We will first formulate a slightly more restrictive set of assumptions which come from Assumptions B. In Section 5.2 we construct the test statistic and remind the general idea of the test. Section 5.3 discusses the appropriate standardization in order to get the desired properties of the test. We demonstrate on the simulations that the usual standardization based on the Fisher information matrix is suitable only in Gaussian case and hence we will come up with the modification which enables to use the test even for distributions which are not Gaussian. Section 5.4 presents the main theorem. Its proof together with the auxiliary lemmas can be found in Section 5.5. The proof goes along the similar steps as in Gombay (2008), but, as told before, we will present it here in a bit more careful way.

We will still consider model (2.1) and SCENARIO 3 as described in Chapter 2, and present slightly more restrictive version of assumptions B:

ASSUMPTIONS B**:

Let condition (B.1) and (B.2)d) be valid and let error term process $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ be a sequence of centered iid random vectors such that

- $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top] = \boldsymbol{\Omega} > 0$,
- $E[\varepsilon_{i,t} \varepsilon_{j,t} \varepsilon_{k,t}] = 0$, $\forall (i, j, k)$ and $\forall t$,
- $E|\varepsilon_{i,t} \varepsilon_{j,t} \varepsilon_{k,t} \varepsilon_{\ell,t}| = \mu_{ijkl} < \infty$, $\forall (i, j, k, \ell)$ and $\forall t$.

Assumptions B** imply Assumptions B. We strengthen the Assumptions B by the independence of error term process $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$. However, Assumptions B** are still weaker than those in Gombay (2008) since we do not assume the Gaussian error term.

We will concentrate on the score test based on the likelihood of a VAR(p) model and derive the asymptotic distribution under H_0 .

5.2 Construction of the test statistic

The test statistic will be derived from the quasi-likelihood function under the assumptions above. Let us denote $\mathbf{K} := \sum_{j=1}^p \Phi_j$, $\mathbf{u}_k := \text{vec}(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_k)$, $\boldsymbol{\mu} := (\mathbf{I}_n - \mathbf{K})^{-1} \mathbf{c}$, $\boldsymbol{\phi} := \text{vec}(\Phi_1, \dots, \Phi_p) \in \mathbb{R}^{n^2 p}$, $\boldsymbol{\omega} := \text{vec}(\boldsymbol{\Omega}) \in \mathbb{R}^{n^2}$, $\boldsymbol{\sigma} := \text{vech}(\boldsymbol{\Omega}) \in \mathbb{R}^{\frac{1}{2}n(n+1)}$, $\boldsymbol{\theta} := (\boldsymbol{\mu}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\sigma}^\top)^\top \in \mathbb{R}^r$, $r := n(np + 1) + \frac{1}{2}n(n + 1)$, $\mathbf{Y}_k := \text{vec}(\mathbf{y}_1, \dots, \mathbf{y}_k) \in \mathbb{R}^{nk}$, $\boldsymbol{\mu}_k := (\boldsymbol{\mu}^\top, \dots, \boldsymbol{\mu}^\top)^\top \in \mathbb{R}^{nk}$ and

$$\mathbf{X}_k = \begin{pmatrix} \mathbf{y}_0 - \boldsymbol{\mu} & \cdots & \mathbf{y}_{k-1} - \boldsymbol{\mu} \\ \vdots & \ddots & \vdots \\ \mathbf{y}_{1-p} - \boldsymbol{\mu} & \cdots & \mathbf{y}_{k-p} - \boldsymbol{\mu} \end{pmatrix} \in \mathbb{R}^{np \times k}.$$

The conditional log-likelihood function ℓ_k based on k observations $\mathbf{y}_1, \dots, \mathbf{y}_k$ with given $\mathbf{y}_{1-p}, \mathbf{y}_{2-p}, \dots, \mathbf{y}_0$ is of the form

$$\ell_k(\boldsymbol{\mu}, \Phi_1, \dots, \Phi_p, \boldsymbol{\Omega}) = -\frac{nk}{2} \ln(2\pi) - \frac{k}{2} \ln |\boldsymbol{\Omega}| - \frac{1}{2} \sum_{t=1}^k \boldsymbol{\varepsilon}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \boldsymbol{\mu} - \sum_{j=1}^p \Phi_j(\mathbf{y}_{t-j} - \boldsymbol{\mu})$. Partial derivatives of ℓ_k with respect to the unknown parameters can be found in Lütkepohl (2005), p. 89–90, and are of the form

$$\begin{aligned} \mathbf{s}_{k1}(\boldsymbol{\theta}) &:= \frac{\partial}{\partial \boldsymbol{\mu}} \ell_k(\boldsymbol{\theta}) = (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^k \left(\mathbf{y}_t - \sum_{j=1}^p \Phi_j \mathbf{y}_{t-j} \right) - \\ &\quad - k(\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} (\mathbf{I}_n - \mathbf{K}) \boldsymbol{\mu} = \\ &= (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^k \boldsymbol{\varepsilon}_t, \\ \mathbf{s}_{k2}(\boldsymbol{\theta}) &:= \frac{\partial}{\partial \boldsymbol{\phi}} \ell_k(\boldsymbol{\theta}) = (\mathbf{X}_k \otimes \boldsymbol{\Omega}^{-1})(\mathbf{Y}_k - \boldsymbol{\mu}_k) - (\mathbf{X}_k \mathbf{X}_k^\top \otimes \boldsymbol{\Omega}^{-1}) \boldsymbol{\phi} = \\ &= (\mathbf{X}_k \otimes \mathbf{I}_n)(\mathbf{I}_k \otimes \boldsymbol{\Omega}^{-1}) \mathbf{u}_k \\ \frac{\partial}{\partial \boldsymbol{\Omega}} \ell_k(\boldsymbol{\theta}) &= -\frac{k}{2} \boldsymbol{\Omega}^{-1} + \frac{1}{2} \boldsymbol{\Omega}^{-1} \cdot \sum_{t=1}^k \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \cdot \boldsymbol{\Omega}^{-1}. \end{aligned}$$

Let us denote $\mathbf{s}_{k3}(\boldsymbol{\theta}) := \text{vech} \left(\frac{\partial}{\partial \boldsymbol{\Omega}} \ell_k(\boldsymbol{\theta}) \right)$ and let $\mathbf{L}_n \in \mathbb{R}^{\frac{1}{2}n(n+1) \times n^2}$ be an elimination matrix such that $\text{vech}(\boldsymbol{\Omega}) = \mathbf{L}_n \text{vec}(\boldsymbol{\Omega})$. Its existence is established in the Appendix A.12 of Lütkepohl (2005). Then

$$\begin{aligned} \mathbf{s}_{k3} &= \mathbf{L}_n \cdot \text{vec} \left(\frac{\partial}{\partial \boldsymbol{\Omega}} \ell_k(\boldsymbol{\theta}) \right) = \frac{1}{2} \mathbf{L}_n \cdot \text{vec} \left(\boldsymbol{\Omega}^{-1} \cdot \sum_{t=1}^k (\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \boldsymbol{\Omega}) \cdot \boldsymbol{\Omega}^{-1} \right) = \\ &= \frac{1}{2} \mathbf{L}_n (\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}) \cdot \sum_{t=1}^k \text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \boldsymbol{\Omega}). \end{aligned}$$

Let $\hat{\boldsymbol{\theta}}_T := (\hat{\boldsymbol{\mu}}_T^\top, \hat{\boldsymbol{\phi}}_T^\top, \hat{\boldsymbol{\sigma}}_T^\top)^\top$ be the maximum likelihood estimators for the unknown set of parameters based on the full sample $\mathbf{y}_1, \dots, \mathbf{y}_T$. For such estimators it

holds $\mathbf{s}_{T1}(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}$, $\mathbf{s}_{T2}(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}$ and $\mathbf{s}_{T3}(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}$. They are of the form

$$\hat{\boldsymbol{\mu}}_T = \frac{1}{T} \left(\mathbf{I}_n - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_{j,T} \right)^{-1} \cdot \sum_{t=1}^T \left(\mathbf{y}_t - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_{j,T} \mathbf{y}_{t-j} \right), \quad (5.1)$$

$$\hat{\boldsymbol{\phi}}_T = \left((\hat{\mathbf{X}}_T \hat{\mathbf{X}}_T^\top)^{-1} \hat{\mathbf{X}}_T \otimes \mathbf{I}_n \right) (\mathbf{Y}_T - \hat{\boldsymbol{\mu}}_T),$$

$$\begin{aligned} \text{vec}(\hat{\boldsymbol{\Omega}}_T) = \hat{\boldsymbol{\omega}}_T &= \frac{1}{T} \text{vec} \left(\sum_{t=1}^T \left(\mathbf{y}_t - \hat{\boldsymbol{\mu}}_T - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_{j,T} \mathbf{y}_{t-j} \right) \cdot \right. \\ &\quad \left. \cdot \left(\mathbf{y}_t - \hat{\boldsymbol{\mu}}_T - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_{j,T} \mathbf{y}_{t-j} \right)^\top \right), \end{aligned} \quad (5.2)$$

where $\hat{\mathbf{X}}_T$ is obtained from \mathbf{X}_T by replacing $\boldsymbol{\mu}$ with the estimate $\hat{\boldsymbol{\mu}}_T$ and $\hat{\boldsymbol{\mu}}_T = \text{vec}(\hat{\boldsymbol{\mu}}_T, \dots, \hat{\boldsymbol{\mu}}_T) \in \mathbb{R}^{nT}$.

The efficient score test statistic is of the following form

$$\hat{\mathbf{B}}(\tau) := \frac{1}{\sqrt{T}} \cdot \mathcal{J}^{-\frac{1}{2}}(\hat{\boldsymbol{\theta}}_T) \cdot \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\mu}} \ell_{[T\tau]}(\hat{\boldsymbol{\theta}}_T) \\ \frac{\partial}{\partial \boldsymbol{\phi}} \ell_{[T\tau]}(\hat{\boldsymbol{\theta}}_T) \\ \frac{\partial}{\partial \boldsymbol{\sigma}} \ell_{[T\tau]}(\hat{\boldsymbol{\theta}}_T) \end{pmatrix}, \quad 0 \leq \tau \leq 1, \quad (5.3)$$

where $[x]$ is integer part of x and \mathcal{J} is a suitable standardization matrix which choice will be discussed in the next section.

The idea behind the test is as follows: It holds that under H_0 , $\mathbf{s}_T(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}$. If H_0 is true, $\mathbf{s}_k(\hat{\boldsymbol{\theta}}_T)$ should be ‘‘close’’ to zero even for the derivatives of the likelihood based on first $k < T$ elements of process $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$. Hence large values of $\sup_{0 \leq \tau \leq 1} \|\hat{\mathbf{B}}(\tau)\|$ will indicate rejection of the null hypothesis.

5.3 Standardization matrix

In case of change detection in univariate AR models, Gombay (2008) suggests using the standardization matrix $\mathcal{I}(\boldsymbol{\theta})$ which is derived from the Fisher information matrix about the parameter $\boldsymbol{\theta}$. It is given by

$$\begin{aligned} \mathcal{I}(\boldsymbol{\theta}) &:= \begin{pmatrix} \mathcal{I}_{1,1}(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{2,2}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}_{3,3}(\boldsymbol{\theta}) \end{pmatrix} = \\ &= \begin{pmatrix} (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} (\mathbf{I}_n - \mathbf{K}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_y(0) \otimes \boldsymbol{\Omega}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{2} \mathbf{D}_n^\top (\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}) \mathbf{D}_n \end{pmatrix}, \end{aligned}$$

see Lütkepohl (2005), p. 91–92, where $\mathbf{D}_n \in \mathbb{R}^{n^2 \times \frac{1}{2}n(n+1)}$ is a duplication matrix such that $\text{vec}(\boldsymbol{\Omega}) = \mathbf{D}_n \text{vech}(\boldsymbol{\Omega})$, $\boldsymbol{\Gamma}_y(0) = \mathbf{E}[\mathbf{Y}_t^{(0)} \mathbf{Y}_t^{(0)\top}]$ and $\mathbf{Y}_t^{(0)\top} := ((\mathbf{y}_{t-1} - \boldsymbol{\mu})^\top, \dots, (\mathbf{y}_{t-p} - \boldsymbol{\mu})^\top)$. Paper by Gombay (2008) considers $\mathcal{I}(\hat{\boldsymbol{\theta}}_T)$, where $\boldsymbol{\Gamma}_y(0)$ is estimated using empirical covariances.

In more general multivariate setting, however, matrix \mathcal{I} is not very suitable since the test statistic is then very sensitive to the underlying distribution. For example, the convergence of the standardized component $\frac{1}{\sqrt{T}} \mathcal{I}^{-\frac{1}{2}}(\hat{\boldsymbol{\theta}}_T) \mathbf{s}_{k3}(\hat{\boldsymbol{\theta}}_T)$ fails

if we consider t -distribution with lower degrees of freedom. The situation is illustrated in Figure 5.1: It depicts the difference between the univariate empirical distribution functions of the components of the test statistic and asymptotic distribution where in both cases matrix $\mathcal{I}(\hat{\boldsymbol{\theta}}_T)$ is chosen. The process is assumed to be P_1 , with $T = 1000$, see Section 3.5 for the details about the simulation settings. We can see very good performance in the left panel where the simulations use the Gaussian distribution for the errors. In case of t_5 -distribution with variance $\frac{5}{3}\boldsymbol{\Omega}$ (right panel), the asymptotic result are satisfactory only for standardized components of \mathbf{s}_{k1} , \mathbf{s}_{k2} , respectively, however convergence fails in case of the component \mathbf{s}_{k3} .

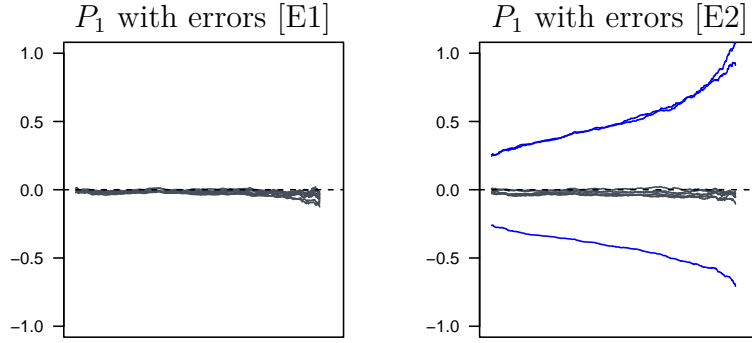


Figure 5.1: Comparison of the asymptotic results. Each grey line represents the difference between the empirical distribution function of the univariate component of the efficient score vector and distribution function of the standard Brownian bridge. Left panel: standard Gaussian errors. Right panel: centered errors with t_5 -distribution and variance $\frac{5}{3}\boldsymbol{\Omega}$. Blue lines in the right panel represent the components of the test statistic belonging to the elements of variance matrix of the error term.

If we consider the theoretical counterpart of the test statistic (5.3) of the form

$$\boldsymbol{\Xi}(\tau) := \frac{1}{\sqrt{T}} \mathcal{J}^{-\frac{1}{2}}(\boldsymbol{\theta}) \mathbf{s}_{\lfloor T\tau \rfloor}(\boldsymbol{\theta}),$$

where $\mathcal{J}(\boldsymbol{\theta}) = \frac{1}{T} \cdot \text{var}[\mathbf{s}_T(\boldsymbol{\theta})]$, then the components of $\boldsymbol{\Xi}(\tau)$ are uncorrelated for all $0 \leq \tau \leq 1$ not only by blocks but also elementwise. Vector $\boldsymbol{\Xi}(\tau)$ is a strong mixing sequence and according to central limit theorem, $\boldsymbol{\Xi}(\tau)$ is asymptotically Gaussian random vector. Hence the components of $\boldsymbol{\Xi}(\tau)$ are asymptotically independent and due to similar arguments and the strong consistency of the estimator $\hat{\boldsymbol{\theta}}_T$ (see Lemma 5.5) the same holds for the test statistic $\hat{\mathbf{B}}(\tau)$. The following Table 5.1 shows the empirical correlations among various components of the test statistic $\hat{\mathbf{B}}$ for different τ and for different distributions of the error term. The notation in the simulation study remains the same as in Section 3.5.

The simulation study reveals the differences in the empirical pairwise correlations among the efficient score components under H_0 . Higher sample correlations can be seen for the case where matrix $\mathcal{I}(\hat{\boldsymbol{\theta}}_T)$ based on Fisher information is selected, especially in the case of t_5 -distribution of the error term. We can also see that the empirical correlation is not affected by the fraction of time τ , where the maximum of the test statistic is achieved.

Table 5.1: Maximal pairwise empirical correlations in absolute value among elements of the test statistic $\widehat{\mathbf{B}}(\tau)$, for different τ and different standardization matrices, and under H_0 .

		Process P_1					
		$T = 100$			$T = 1000$		
		$\tau = \frac{1}{10}$	$\tau = \frac{1}{2}$	$\tau = \frac{9}{10}$	$\tau = \frac{1}{10}$	$\tau = \frac{1}{2}$	$\tau = \frac{9}{10}$
[E1]	(\mathcal{J})	0.09	0.10	0.07	0.06	0.09	0.06
[E1]	(\mathcal{I})	0.08	0.10	0.10	0.06	0.06	0.08
[E2]	(\mathcal{J})	0.09	0.09	0.07	0.08	0.06	0.07
[E2]	(\mathcal{I})	0.18	0.20	0.22	0.28	0.33	0.35

Going back to the standardization matrix \mathcal{J} it is not difficult to show that under Assumptions B**

$$\mathcal{J}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{J}_{1,1}(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{J}_{2,2}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{J}_{3,3}(\boldsymbol{\theta}) \end{pmatrix},$$

where

$$\begin{aligned} \mathcal{J}_{1,1}(\boldsymbol{\theta}) &= (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} (\mathbf{I}_n - \mathbf{K}), \\ \mathcal{J}_{2,2}(\boldsymbol{\theta}) &= \boldsymbol{\Gamma}_y(0) \otimes \boldsymbol{\Omega}^{-1}, \\ \mathcal{J}_{3,3}(\boldsymbol{\theta}) &= \frac{1}{4} \mathbf{L}_n (\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}) \mathbf{V} (\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}) \mathbf{L}_n^\top, \end{aligned}$$

where $\mathcal{J}_{i,i}(\boldsymbol{\theta}) = \frac{1}{T} \text{var}[\mathbf{s}_{Ti}(\boldsymbol{\theta})]$, $i = 1, \dots, 3$, $\mathbf{V} = \text{var}[\text{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top - \boldsymbol{\Omega})]$. Comparing the diagonal elements of the standardization matrices \mathcal{I} and \mathcal{J} we see that $\mathcal{I}_{i,i} = \mathcal{J}_{i,i}$, $i = 1, 2$, but generally $\mathcal{I}_{3,3} \neq \mathcal{J}_{3,3}$. Relation $\mathcal{I}_{3,3} = \mathcal{J}_{3,3}$ holds in Gaussian case.

In order to use the matrix \mathcal{J} in the test statistic, it is necessary to find an estimate $\widehat{\mathbf{V}}_T$ for \mathbf{V} such that $\widehat{\mathbf{V}}_T - \mathbf{V} = o_P(1)$, as $T \rightarrow \infty$. A convenient estimator is the empirical counterpart of \mathbf{V} of the form

$$\widehat{\mathbf{V}}_T = \frac{1}{T} \cdot \sum_{t=1}^T \left(\text{vec}(\widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t^\top) \text{vec}(\widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t^\top)^\top - \widehat{\boldsymbol{\omega}}_T \widehat{\boldsymbol{\omega}}_T^\top \right).$$

In what follows, let $\mathcal{J}(\widehat{\boldsymbol{\theta}}_T)$ be the standardization matrix, where the true parameter $\boldsymbol{\theta}$ is replaced by $\widehat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\Gamma}_y(0)$ and \mathbf{V} are replaced by their sample counterparts.

Throughout the rest of the chapter, we will omit the subscript T in the ML estimators for notation simplicity. Let $\boldsymbol{\theta}$ be the vector with the true values of the parameters. Let us denote $\widehat{\mathbf{K}} = \sum_{j=1}^p \widehat{\boldsymbol{\Phi}}_j$.

5.4 Main result

The aim of this chapter will be the proof of the following theorem:

Theorem 5.1 *Let us suppose that the sequence $\{\mathbf{y}_t\}_{t \in \mathbb{Z}}$ follows VAR(p) model and Assumptions B** are fulfilled. Then, under H_0 , there exists an r -dimensional*

sequence of standard Brownian bridges $\mathbf{B}(\tau)$ with independent components $B_j(\tau)$, $0 \leq \tau \leq 1$, $j = 1, \dots, r$, such that

$$\max_{j=1, \dots, r} \sup_{0 \leq \tau \leq 1} |\widehat{B}_j(\tau) - B_j(\tau)| = o_{\mathbb{P}}(1), \quad T \rightarrow \infty. \quad (5.4)$$

Let us construct the test statistic based on the Theorem 5.1. The asymptotic result (5.4) implies that under H_0

$$\sup_{0 \leq \tau \leq 1} |\widehat{B}_j(\tau)| \xrightarrow{\mathbb{P}} \sup_{0 \leq \tau \leq 1} |B_j(\tau)| \quad j = 1, \dots, r.$$

If we want to test a change in one arbitrary parameter of interest, say j , we reject H_0 if

$$\sup_{0 \leq \tau \leq 1} |\widehat{B}_j(\tau)| \geq C(\alpha),$$

and the critical value $C(\alpha)$, as pointed out in Gombay (2008), can be obtained from the relationship

$$\mathbb{P} \left[\sup_{0 \leq \tau \leq 1} |B_1(\tau)| > x \right] = 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 x^2}. \quad (5.5)$$

If we want to test a change in d parameters, $1 \leq d \leq r$, and keep the significance level α , then

$$\begin{aligned} \alpha &= \mathbb{P}_{H_0} \left[\max_{j=1, \dots, d} \sup_{0 \leq \tau \leq 1} |B_j(\tau)| > C(\alpha) \right] = \\ &= 1 - \mathbb{P}_{H_0} \left[\sup_{0 \leq \tau \leq 1} |B_j(\tau)| \leq C(\alpha), \forall j = 1, \dots, d \right] = \\ &= 1 - \prod_{j=1}^d \mathbb{P}_{H_0} \left[\sup_{0 \leq \tau \leq 1} |B_j(\tau)| \leq C(\alpha) \right] = \\ &= 1 - \prod_{j=1}^d \left(1 - \mathbb{P}_{H_0} \left[\sup_{0 \leq \tau \leq 1} |B_j(\tau)| > C(\alpha) \right] \right) = \\ &= 1 - \left(1 - \mathbb{P}_{H_0} \left[\sup_{0 \leq \tau \leq 1} |B_1(\tau)| > C(\alpha) \right] \right)^d = 1 - (1 - \alpha^*)^d, \end{aligned}$$

from which we obtain the ‘‘individual’’ level $\alpha^* = 1 - (1 - \alpha)^{\frac{1}{d}}$.

Table 5.2 shows the critical values based on the approximation (5.5) for different number of parameters subject to a change. This value is compared to the value obtained in a Monte Carlo simulation study where we simulate $\max_{j=1, \dots, d} \sup_{0 \leq \tau \leq 1} |B_j(\tau)|$ for different $d = 1, \dots, 9$, 10 000-times with a grid of 10 000 for a Brownian bridge process.

Table 5.2 serves only as a comparison. In the section which contains the simulation study, the critical values are based on the exact result in (5.5).

5.5 Proofs

A multivariate form of the Theorem 1.2.1 of Csörgő and Révész (1981) will be proven first:

Table 5.2: Critical values based on (5.5) and Monte Carlo (MC) simulations for different levels α and various number of parameters $d = 1, \dots, 9$ which are subject to a change.

d	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
	(5.5)	MC	(5.5)	MC	(5.5)	MC
1	1.224	1.213	1.358	1.360	1.628	1.625
2	1.353	1.344	1.478	1.470	1.730	1.735
3	1.425	1.419	1.544	1.522	1.788	1.777
4	1.474	1.470	1.590	1.577	1.828	1.817
5	1.511	1.501	1.624	1.611	1.858	1.845
6	1.540	1.540	1.652	1.655	1.882	1.884
7	1.565	1.566	1.675	1.668	1.903	1.888
8	1.586	1.579	1.695	1.679	1.920	1.897
9	1.604	1.595	1.712	1.708	1.935	1.906

Theorem 5.2 *If \mathbf{W} is a d -dimensional standard Wiener process, it holds*

$$\limsup_{k \rightarrow \infty} \sup_{0 \leq s \leq p} \|\mathbf{W}(k-s) - \mathbf{W}(k)\| = \mathcal{O}(\sqrt{\ln k}) \quad \text{a.s.}$$

Proof: According to the Theorem 1.2.1 of Csörgő and Révész: If a_T is monotonically non-decreasing function of T , $0 < a_T \leq T$ and $\frac{T}{a_T}$ is monotonically non-decreasing, then for univariate standard Wiener process W

$$\limsup_{T \rightarrow \infty} \sup_{0 \leq s \leq a_T} \beta_T |W(T+s) - W(T)| = 1, \quad \text{a.s.}, \quad (5.6)$$

where $\beta_T = \left(2a_T \left(\ln \frac{T}{a_T} + \ln \ln T\right)\right)^{-\frac{1}{2}}$. If we substitute $T := k-s$, $a_T := p$ into (5.6), we get that

$$\limsup_{k \rightarrow \infty} \sup_{0 \leq s \leq p} |W(k-s) - W(k)| = \mathcal{O}(\sqrt{\ln k}) \quad \text{a.s.}, \quad T \rightarrow \infty.$$

Now, if $\mathbf{W}(t) = (W_1(t), \dots, W_d(t))^\top$ then

$$\begin{aligned} & \limsup_{k \rightarrow \infty} \sup_{0 \leq s \leq p} \|\mathbf{W}(k-s) - \mathbf{W}(k)\| = \\ &= \limsup_{k \rightarrow \infty} \sup_{0 \leq s \leq p} \sqrt{\sum_{j=1}^d [W_j(k-s) - W_j(k)]^2} \leq \\ &\leq \sum_{j=1}^d \limsup_{k \rightarrow \infty} \sup_{0 \leq s \leq p} |W_j(k-s) - W_j(k)| = \mathcal{O}(\sqrt{\ln k}) \quad \text{a.s.}, \quad k \rightarrow \infty, \end{aligned}$$

which concludes the proof. □

The following theorem is FLCT for the VAR(p) stationary process.

Theorem 5.3 *Under Assumptions B^{**} it holds that*

$$\left\| \sum_{t=1}^{\lfloor \tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(\tau) \right\| = o\left(\tau^{\frac{1}{\nu}}\right) \quad \text{a.s.}, \quad (5.7)$$

5. SCORE TEST

for some $\nu > 2$ and some $\Sigma > 0$, where $\mathbf{W}(\tau)$ is an n -dimensional standard Wiener process.

Proof: It follows immediately from FCLT B.10, since under iid errors and other assumptions in B^{**} , sequence $\{\mathbf{y}_t - \boldsymbol{\mu}\}_{t \in \mathbb{Z}}$ is weakly stationary centered strong mixing sequence with uniformly bounded $(2 + \delta)$ -moment, for some $\delta > 0$. As regards the rate of convergence, $\mathcal{O}(\tau^{\frac{1}{2} - \delta})$ a.s. implies $o(\tau^{\frac{1}{2} - \frac{\delta}{2(2+\delta)}})$ a.s., for some $\delta > 0$, which is $o(\tau^{\frac{1}{\nu}})$ a.s., for some $\nu > 2$. \square

Note: The preceding theorem, as well as the following lemma, is stated for sequence $\{\mathbf{y}_t - \boldsymbol{\mu}\}_{t \in \mathbb{Z}}$, however the statements would be valid for any sequence of random vectors fulfilling conditions of FCLT B.10.

Lemma 5.4 *Under Assumptions B^{**} it holds that*

$$\sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \cdot \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \Sigma \mathbf{W}(\tau) \right\| = o_{\mathbb{P}}(1), \quad T \rightarrow \infty,$$

for some $\Sigma > 0$, where $\mathbf{W}(\tau)$ is an n -dimensional standard Wiener process

Proof: From Theorem (5.3)

$$\frac{1}{\sqrt{s}} \left\| \sum_{t=1}^s (\mathbf{y}_t - \boldsymbol{\mu}) - \Sigma \mathbf{W}(s) \right\| = o(1) \quad \text{a.s.}, \quad s \rightarrow \infty.$$

Let $s := \lfloor T\tau \rfloor$, $0 \leq \tau \leq 1$. Then

$$\frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \Sigma \mathbf{W}(T\tau) \right\| = o(1) \quad \text{a.s.}, \quad T \rightarrow \infty, \quad (5.8)$$

and hence

$$\begin{aligned} & \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \Sigma \mathbf{W}(\tau) \right\| = \\ &= \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \frac{1}{\sqrt{T}} \Sigma \mathbf{W}(T\tau) + \frac{1}{\sqrt{T}} \Sigma \mathbf{W}(T\tau) - \Sigma \mathbf{W}(\tau) \right\| \leq \\ &\leq \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \Sigma \mathbf{W}(T\tau) \right\| + \|\Sigma\| \cdot \left\| \frac{1}{\sqrt{T}} \mathbf{W}(T\tau) - \mathbf{W}(\tau) \right\|. \end{aligned}$$

The first addend is $o(1)$, as $T \rightarrow \infty$, due to (5.8). For the second addend we use the well known fact that $\frac{1}{\sqrt{T}} \mathbf{W}(T\tau) \stackrel{d}{=} \mathbf{W}(\tau)$ from which follows that $\frac{1}{\sqrt{T}} \mathbf{W}(T\tau) - \mathbf{W}(\tau) = o_{\mathbb{P}}(1)$, $T \rightarrow \infty$. The rest of the proof follows from Continuous Mapping Theorem. \square

Lemma 5.5 *Estimators $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\Omega}}$ of the VAR(p) model under the Assumptions B^{**} fulfill*

$$(a) \quad \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| = \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \quad \text{a.s.}, \quad (5.9)$$

$$(b) \quad \|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\| = \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \quad \text{a.s.}, \quad (5.10)$$

$$(c) \quad \|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\| = \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \quad \text{a.s.} \quad (5.11)$$

Proof: It follows immediately from Theorem 3.8. \square

Lemma 5.6 *Under Assumptions B^{**} it holds*

$$\frac{1}{\sqrt{T}} \cdot \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} (\hat{\boldsymbol{\varepsilon}}_t - \boldsymbol{\varepsilon}_t) \right\| = \mathcal{O}(\sqrt{\ln \ln T}) \quad \text{a.s.}, \quad T \rightarrow \infty,$$

where $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\boldsymbol{\mu}} - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \hat{\boldsymbol{\mu}})$.

Proof: Before analyzing the convergence, let us expand the difference between the residuals and errors as follows:

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}}_t - \boldsymbol{\varepsilon}_t &= \boldsymbol{\mu} - \hat{\boldsymbol{\mu}} - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \boldsymbol{\mu} - \hat{\boldsymbol{\mu}} + \boldsymbol{\mu}) + \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{y}_{t-j} - \boldsymbol{\mu}) = \\ &= \boldsymbol{\mu} - \hat{\boldsymbol{\mu}} - \sum_{j=1}^p (\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) (\mathbf{y}_{t-j} - \boldsymbol{\mu}) - \sum_{j=1}^p (\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) - \\ &\quad - \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{y}_{t-j} - \boldsymbol{\mu}) = -(\mathbf{I}_n - \mathbf{K})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) - \\ &\quad - \sum_{j=1}^p (\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) (\mathbf{y}_{t-j} - \boldsymbol{\mu}) + \sum_{j=1}^p (\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}). \end{aligned} \quad (5.12)$$

Let M be a generic constant. Using the above expansion,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} (\hat{\boldsymbol{\varepsilon}}_t - \boldsymbol{\varepsilon}_t) \right\| &= \frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \left\| -\lfloor T\tau \rfloor \cdot (\mathbf{I}_n - \mathbf{K})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) - \right. \\ &\quad \left. - \sum_{j=1}^p \left[(\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) \left(\sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_{t-j} - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T\tau - j) \right) \right] - \right. \\ &\quad \left. - \sum_{j=1}^p (\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) \boldsymbol{\Sigma} \mathbf{W}(T\tau - j) + \lfloor T\tau \rfloor \sum_{j=1}^p (\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \right\| \leq \\ &\leq M\sqrt{T} \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| + p \cdot \max_{1 \leq j \leq p} \|\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j\| \cdot \\ &\quad \frac{1}{\sqrt{T}} \max_{1 \leq j \leq p} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_{t-j} - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T\tau - j) \right\| + \\ &\quad + M \max_{1 \leq j \leq p} \|\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j\| \cdot \frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \|\mathbf{W}(T\tau - j)\| + \\ &\quad + \sqrt{T} \max_{1 \leq j \leq p} \|\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j\| \cdot \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|, \end{aligned}$$

5. SCORE TEST

for some $\Sigma > 0$.

Since, as $T \rightarrow \infty$,

$$\begin{aligned} & \frac{1}{\sqrt{T}} \max_{1 \leq j \leq p} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_{t-j} - \boldsymbol{\mu}) - \Sigma \mathbf{W}(T\tau - j) \right\| = \\ & = \frac{1}{\sqrt{T}} \max_{1 \leq j \leq p} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor - j} (\mathbf{y}_t - \boldsymbol{\mu}) - \Sigma \mathbf{W}(T\tau - j) \right\| + o(1) \quad \text{a.s.}, \end{aligned} \quad (5.13)$$

the rest of the proof follows from Lemmas 5.4 and 5.5 \square

The next lemma says that it is possible to interchange the sample mean and the ML estimate in VAR(p) model. As stated for instance in Lütkepohl (2005), p. 92, or Proposition 3.3, the ML estimate is asymptotically equivalent to the sample mean, but the rate of the asymptotic difference between those two is not published. However, Gombay (2008) explicitly mention this rate but the reason is rather short and gives no insight. We bring a more detailed proof of the statement of Gombay (2008) in case of a VAR(p) model:

Lemma 5.7 *If $\hat{\boldsymbol{\mu}}$ is the ML estimate and $\bar{\mathbf{y}} := \bar{\mathbf{y}}_T = T^{-1} \sum_{t=1}^T \mathbf{y}_t$ is the sample mean, then under conditions B^{**} it holds $\|\hat{\boldsymbol{\mu}} - \bar{\mathbf{y}}\| = o(T^{\frac{1}{\nu}-1})$ a.s., for some $\nu > 2$.*

Proof: By plugging in $\bar{\mathbf{y}}$ inside both sums in expression (5.1),

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{T} (\mathbf{I}_n - \hat{\mathbf{K}})^{-1} \sum_{t=1}^T (\mathbf{y}_t - \bar{\mathbf{y}} - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}})) + \\ & \quad + \frac{1}{T} (\mathbf{I}_n - \hat{\mathbf{K}})^{-1} T \bar{\mathbf{y}} - \frac{1}{T} (\mathbf{I}_n - \hat{\mathbf{K}})^{-1} \sum_{t=1}^T \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j \bar{\mathbf{y}} = \\ &= \frac{1}{T} (\mathbf{I}_n - \hat{\mathbf{K}})^{-1} \sum_{t=1}^T (\mathbf{y}_t - \bar{\mathbf{y}} - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}})) + \bar{\mathbf{y}}. \end{aligned}$$

Hence

$$\begin{aligned} \hat{\boldsymbol{\mu}} - \bar{\mathbf{y}} &= \frac{1}{T} (\mathbf{I}_n - \hat{\mathbf{K}})^{-1} \sum_{t=1}^T (\mathbf{y}_t - \bar{\mathbf{y}} - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}})) = \\ &= \frac{1}{T} (\mathbf{I}_n - \hat{\mathbf{K}})^{-1} \sum_{t=1}^T (\mathbf{y}_t - \bar{\mathbf{y}}) - \frac{1}{T} (\mathbf{I}_n - \hat{\mathbf{K}})^{-1} \sum_{t=1}^T \left(\sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right) = \\ &= -\frac{1}{T} (\mathbf{I}_n - \hat{\mathbf{K}})^{-1} \sum_{t=1}^T \left(\sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right). \end{aligned}$$

Since

$$\begin{aligned} \|\hat{\boldsymbol{\mu}} - \bar{\mathbf{y}}\| &\leq \underbrace{\left\| \frac{1}{T} [(\mathbf{I}_n - \hat{\mathbf{K}})^{-1} - (\mathbf{I}_n - \mathbf{K})^{-1}] \cdot \sum_{t=1}^T \left(\sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right) \right\|}_{(\text{CONV1})} + \\ & \quad + \underbrace{\left\| \frac{1}{T} (\mathbf{I}_n - \mathbf{K})^{-1} \sum_{t=1}^T \left(\sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right) \right\|}_{(\text{CONV2})} \end{aligned}$$

and (CONV1) will achieve better convergence rate than (CONV2) due to Theorem 5.5, it suffices to treat (CONV2):

$$\begin{aligned}
 & \frac{1}{T} \left\| (\mathbf{I}_n - \mathbf{K})^{-1} \sum_{t=1}^T \left(\sum_{j=1}^p \hat{\Phi}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right) \right\| \leq M \cdot \\
 & \cdot \frac{1}{T} \left\| \sum_{j=1}^p \left[(\hat{\Phi}_j - \Phi_j) \left(\sum_{t=1}^T (\mathbf{y}_{t-j} - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T-j) \right) \right] + \right. \\
 & + \sum_{j=1}^p (\hat{\Phi}_j - \Phi_j) \boldsymbol{\Sigma} \mathbf{W}(T-j) - \\
 & - \sum_{j=1}^p \left[(\hat{\Phi}_j - \Phi_j) \left(\sum_{s=1}^T (\mathbf{y}_s - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T) \right) \right] - \sum_{j=1}^p (\hat{\Phi}_j - \Phi_j) \boldsymbol{\Sigma} \mathbf{W}(T) + \\
 & + \sum_{j=1}^p \left[\Phi_j \left(\sum_{t=1}^T (\mathbf{y}_{t-j} - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T-j) \right) \right] + \sum_{j=1}^p \Phi_j \boldsymbol{\Sigma} \mathbf{W}(T-j) - \\
 & \left. - \sum_{j=1}^p \left[\Phi_j \left(\sum_{s=1}^T (\mathbf{y}_s - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T) \right) \right] - \sum_{j=1}^p \Phi_j \boldsymbol{\Sigma} \mathbf{W}(T) \right\| \leq \\
 \leq & Mp \cdot \max_{1 \leq j \leq p} \|\hat{\Phi}_j - \Phi_j\| \cdot \frac{1}{T} \max_{1 \leq j \leq p} \left\| \sum_{t=1}^T (\mathbf{y}_{t-j} - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T-j) \right\| + \\
 & + Mp \cdot \max_{1 \leq j \leq p} \|\hat{\Phi}_j - \Phi_j\| \cdot \frac{1}{T} \|\boldsymbol{\Sigma}\| \max_{1 \leq j \leq p} \|\mathbf{W}(T-j) - \mathbf{W}(T)\| + \\
 & + Mp \cdot \max_{1 \leq j \leq p} \|\hat{\Phi}_j - \Phi_j\| \cdot \frac{1}{T} \left\| \sum_{s=1}^T (\mathbf{y}_s - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T) \right\| + \\
 & + Mp \cdot \max_{1 \leq j \leq p} \|\Phi_j\| \cdot \frac{1}{T} \max_{1 \leq j \leq p} \left\| \sum_{t=1}^T (\mathbf{y}_{t-j} - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T-j) \right\| + \\
 & + Mp \cdot \max_{1 \leq j \leq p} \|\Phi_j\| \cdot \frac{1}{T} \|\boldsymbol{\Sigma}\| \max_{1 \leq j \leq p} \|\mathbf{W}(T-j) - \mathbf{W}(T)\| + \\
 & + Mp \cdot \max_{1 \leq j \leq p} \|\Phi_j\| \cdot \frac{1}{T} \left\| \sum_{s=1}^T (\mathbf{y}_s - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(T) \right\| =: \\
 =: & a_T^{(1)} + a_T^{(2)} + a_T^{(3)} + a_T^{(4)} + a_T^{(5)} + a_T^{(6)},
 \end{aligned}$$

where $M = \|(\mathbf{I}_n - \mathbf{K})^{-1}\|$, for some $\boldsymbol{\Sigma} > 0$. Now, due to (5.13), Lemma 5.5 and Theorem 5.3, $a_T^{(1)}$ and $a_T^{(3)}$ are $o(T^{\frac{1}{\nu} - \frac{3}{2}} \sqrt{\ln \ln T})$, a.s., $a_T^{(4)}$ and $a_T^{(6)}$ are $o(T^{\frac{1}{\nu} - 1})$, a.s., as $T \rightarrow \infty$, for some $\nu > 2$. Due to Theorem 5.2 and Lemma 5.5, $a_T^{(2)}$ is $\mathcal{O}(T^{-\frac{3}{2}} \sqrt{\ln T \ln \ln T})$, a.s., $a_T^{(5)}$ is $\mathcal{O}(T^{-1} \sqrt{\ln T})$, a.s., as $T \rightarrow \infty$. Hence the worst overall rate is $o(T^{\frac{1}{\nu} - 1})$, a.s., as $T \rightarrow \infty$, for some $\nu > 2$. \square

Proof of Theorem 5.1: The proof will be divided into three parts as in Gombay (2008). Because ML estimators are $\mathcal{O}_P(T^{-\frac{1}{2}})$, see Lütkepohl (2005) for details, it follows that $\|\mathcal{J}(\hat{\boldsymbol{\theta}}) - \mathcal{J}(\boldsymbol{\theta})\| = o_P(1)$, as $T \rightarrow \infty$. We are going to show that $T^{-\frac{1}{2}}$ times all the partial derivatives of the score vector are $o_P(1)$ and this will conclude the proof.

5. SCORE TEST

(i) Let us consider the change in $\boldsymbol{\mu}$. The component of the efficient score vector that we use is

$$\begin{aligned} & \frac{1}{\sqrt{T}} \left[\frac{\partial}{\partial \boldsymbol{\mu}} \ell_k(\boldsymbol{\theta}) \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \\ & = \frac{1}{\sqrt{T}} \left(\mathbf{I}_n - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j \right)^\top \hat{\boldsymbol{\Omega}}^{-1} \sum_{t=1}^k \left(\mathbf{y}_t - \hat{\boldsymbol{\mu}} - \sum_{j=1}^p \hat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \hat{\boldsymbol{\mu}}) \right) + \end{aligned} \quad (5.14)$$

$$+ \frac{1}{\sqrt{T}} \left(\mathbf{I}_n - \sum_{j=1}^p \boldsymbol{\Phi}_j \right)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^k \left(\mathbf{y}_t - \bar{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right) - \quad (5.15)$$

$$- \frac{1}{\sqrt{T}} \left(\mathbf{I}_n - \sum_{j=1}^p \boldsymbol{\Phi}_j \right)^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^k \left(\mathbf{y}_t - \bar{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right). \quad (5.16)$$

We will show

- (1) the convergence of the supremum of the norm of the second addend (5.15) to the supremum of the norm of Brownian bridge process, and
- (2) also that the supremum of the norm of the difference between (5.14) and (5.16) converges to zero in probability,

which will complete the first part of the proof.

(1) Let $k = \lfloor T\tau \rfloor$, $0 \leq \tau \leq 1$. In this part of the proof, M is a positive constant which can be different from term to term. For the second addend (5.15) we obtain

$$\begin{aligned} & \sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^{\lfloor T\tau \rfloor} \left(\mathbf{y}_t - \bar{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right) - \mathbf{B}(\tau) \right\| \leq \\ & \leq M \sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \left(\sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \bar{\mathbf{y}}) - \sum_{j=1}^p \boldsymbol{\Phi}_j \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right) - \boldsymbol{\Sigma} \mathbf{B}(\tau) \right\|, \end{aligned}$$

for some $\boldsymbol{\Sigma} > 0$. The last term is equal to

$$\begin{aligned} & \sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \bar{\mathbf{y}}) - \boldsymbol{\Sigma} \mathbf{B}(\tau) - \frac{1}{\sqrt{T}} \sum_{j=1}^p \boldsymbol{\Phi}_j \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right\| \leq \\ & \leq \sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \bar{\mathbf{y}}) - \boldsymbol{\Sigma} \mathbf{B}(\tau) \right\| + \sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \sum_{j=1}^p \boldsymbol{\Phi}_j \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right\| = \\ & =: A_T + B_T. \end{aligned}$$

Let us focus on the term A_T . Because $\mathbf{B}(\tau) = \mathbf{W}(\tau) - \tau \mathbf{W}(1)$, it holds that

$$\begin{aligned} A_T & = \sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \bar{\mathbf{y}}) - \boldsymbol{\Sigma} \mathbf{W}(\tau) + \tau \boldsymbol{\Sigma} \mathbf{W}(1) \right\| = \\ & = \sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \frac{\lfloor T\tau \rfloor}{T\sqrt{T}} \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(\tau) + \boldsymbol{\Sigma} \tau \mathbf{W}(1) \right\| \leq \\ & \leq \sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(\tau) \right\| + \\ & \quad + \sup_{0 \leq \tau \leq 1} \left\{ |\tau| \cdot \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\mu}) - \boldsymbol{\Sigma} \mathbf{W}(1) \right\| \right\}. \end{aligned}$$

5. SCORE TEST

Both latter terms are $o_{\mathbb{P}}(1)$, as $T \rightarrow \infty$, according to Lemma 5.4. Hence $A_T = o_{\mathbb{P}}(1)$, as $T \rightarrow \infty$. Now let us focus on B_T . Convergence remains unaffected when replacing $\bar{\mathbf{y}}$ with $\bar{\mathbf{y}}_k = k^{-1} \sum_{t=1}^k \mathbf{y}_t$. Hence

$$\begin{aligned} B_T \leq & M \cdot \left[\frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \sup_{1 \leq j \leq p} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor - j} (\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{W}(T\tau - j) \right\| + \right. \\ & + \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_t - \boldsymbol{\mu}) - \mathbf{W}(T\tau) \right\| + \\ & \left. + \frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \sup_{1 \leq j \leq p} \left\| \mathbf{W}(T\tau - j) - \mathbf{W}(T\tau) \right\| \right] = o_{\mathbb{P}}(1), \quad T \rightarrow \infty \end{aligned}$$

due to argument (5.13), Theorem 5.3 and Theorem 5.2.

(2) Now let us focus on the supremum over $0 \leq \tau \leq 1$ of the norm of the difference between (5.14) and (5.16):

$$\begin{aligned} & \frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \left\| (\mathbf{I}_n - \widehat{\mathbf{K}})^\top \widehat{\boldsymbol{\Omega}}^{-1} \sum_{t=1}^{\lfloor T\tau \rfloor} \widehat{\boldsymbol{\varepsilon}}_t - (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^{\lfloor T\tau \rfloor} \widehat{\boldsymbol{\varepsilon}}_t + \right. \\ & + \left. (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^{\lfloor T\tau \rfloor} \widehat{\boldsymbol{\varepsilon}}_t - (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} \sum_{t=1}^{\lfloor T\tau \rfloor} \left(\mathbf{y}_t - \bar{\mathbf{y}} - \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right) \right\| \leq \\ & \leq \left\| (\mathbf{I}_n - \widehat{\mathbf{K}})^\top \widehat{\boldsymbol{\Omega}}^{-1} - (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} \right\| \cdot \frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} \widehat{\boldsymbol{\varepsilon}}_t \right\| + \end{aligned} \quad (5.17)$$

$$+ M \cdot \frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} \left[\widehat{\boldsymbol{\varepsilon}}_t - \mathbf{y}_t + \bar{\mathbf{y}} + \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right] \right\|, \quad (5.18)$$

where, in this case, $M = \|(\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1}\|$. Let us focus on the matrices in the first norm in (5.17):

$$\begin{aligned} (\mathbf{I}_n - \widehat{\mathbf{K}})^\top \widehat{\boldsymbol{\Omega}}^{-1} - (\mathbf{I}_n - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} &= (\mathbf{I}_n - \mathbf{K})^\top (\widehat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1}) - \\ & - (\widehat{\mathbf{K}} - \mathbf{K})^\top \boldsymbol{\Omega}^{-1} - (\widehat{\mathbf{K}} - \mathbf{K})^\top (\widehat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1}) = \\ & = \mathcal{O}\left(\sqrt{\frac{\ln \ln T}{T}}\right) \quad \text{a.s., } T \rightarrow \infty, \end{aligned}$$

due to Theorem 5.5. Now,

$$\frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} \widehat{\boldsymbol{\varepsilon}}_t \right\| \leq \underbrace{\frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} (\widehat{\boldsymbol{\varepsilon}}_t - \boldsymbol{\varepsilon}_t) \right\|}_{(\text{ADD1})} + \underbrace{\frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} \boldsymbol{\varepsilon}_t \right\|}_{(\text{ADD2})}.$$

Due to Lemma 5.6 the first addend (ADD1) is $\mathcal{O}(\sqrt{\ln \ln T})$ a.s., as $T \rightarrow \infty$, and for (ADD2) we get the same rate by using the law of the iterated logarithm for iid vectors. Hence, (5.17) is $\mathcal{O}(T^{-\frac{1}{2}} \ln \ln T)$ a.s., as $T \rightarrow \infty$.

5. SCORE TEST

For term (5.18), by using $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \widehat{\boldsymbol{\mu}} - \sum_{j=1}^p \widehat{\boldsymbol{\Phi}}_j (\mathbf{y}_{t-j} - \widehat{\boldsymbol{\mu}})$, we get that (5.18) is less or equal to

$$\begin{aligned}
& M \frac{1}{\sqrt{T}} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} \left[(\mathbf{I}_n - \mathbf{K})(\bar{\mathbf{y}} - \widehat{\boldsymbol{\mu}}) - \sum_{j=1}^p (\widehat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j)(\mathbf{y}_{t-j} - \bar{\mathbf{y}}) + \right. \right. \\
& \quad \left. \left. + \sum_{j=1}^p (\widehat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j)(\widehat{\boldsymbol{\mu}} - \bar{\mathbf{y}}) \right] \right\| \leq \\
& \leq M \cdot \left[\sqrt{T} \|\widehat{\boldsymbol{\mu}} - \bar{\mathbf{y}}\| + \max_{1 \leq j \leq p} \|\widehat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j\| \cdot \frac{1}{\sqrt{T}} \max_{1 \leq j \leq p} \sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} (\mathbf{y}_{t-j} - \bar{\mathbf{y}}) \right\| + \right. \\
& \quad \left. + \sqrt{T} \max_{1 \leq j \leq p} \|\widehat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j\| \cdot \|\widehat{\boldsymbol{\mu}} - \bar{\mathbf{y}}\| \right] =: \\
& =: \eta_T^{(1)} + \eta_T^{(2)} + \eta_T^{(3)}.
\end{aligned}$$

According to Lemma 5.7 it holds that $\eta_T^{(1)} = o(T^{\frac{1}{\nu} - \frac{1}{2}})$ a.s., for some $\nu > 2$. Term $\eta_T^{(2)} = \mathcal{O}_P((T^{-1} \ln \ln T)^{\frac{1}{2}})$ a.s., $T \rightarrow \infty$, due to Lemma 5.5 and due to the arguments in the proof of convergence for B_T from step (1). Term $\eta_T^{(3)} = o(T^{\frac{1}{\nu} - 1} \sqrt{\ln \ln T})$ a.s., for some $\nu > 2$, $T \rightarrow \infty$. Hence the proof for part (i) is completed.

(ii) Let us suppose we want to test changes in $\boldsymbol{\Phi}_s$, $s = 1, \dots, p$. First, let us denote $\mathbf{v}_t := \text{vec}(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}) \in \mathbb{R}^{np}$, $\boldsymbol{\mu}_p := \text{vec}(\boldsymbol{\mu}, \dots, \boldsymbol{\mu}) \in \mathbb{R}^{np}$, $\mathbf{M}_t := (\mathbf{v}_t - \boldsymbol{\mu}_p)^\top \otimes \mathbf{I}_n \in \mathbb{R}^{n \times n^2 p}$ and $\widehat{\mathbf{M}}_t := (\mathbf{v}_t - \widehat{\boldsymbol{\mu}}_p)^\top \otimes \mathbf{I}_n$ be the estimate of \mathbf{M}_t . Then

$$\begin{aligned}
\left[\frac{\partial}{\partial \boldsymbol{\phi}} \ell_k(\boldsymbol{\theta}) \right]_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}} &= (\widehat{\mathbf{X}}_k \otimes \widehat{\boldsymbol{\Omega}}^{-1})(\mathbf{Y}_k - \widehat{\boldsymbol{\mu}}_k) - ((\widehat{\mathbf{X}}_k \widehat{\mathbf{X}}_k^\top) \otimes \widehat{\boldsymbol{\Omega}}^{-1}) \widehat{\boldsymbol{\phi}} = \\
&= \sum_{t=1}^k \widehat{\mathbf{M}}_t^\top \widehat{\boldsymbol{\Omega}}^{-1} (\mathbf{y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\mathbf{M}}_t \widehat{\boldsymbol{\phi}}) = \sum_{t=1}^k \widehat{\mathbf{M}}_t^\top \widehat{\boldsymbol{\Omega}}^{-1} \widehat{\boldsymbol{\varepsilon}}_t.
\end{aligned}$$

Replacing $\widehat{\boldsymbol{\Omega}}$ with $\boldsymbol{\Omega}$ does not change the asymptotic distribution. By Lemma 5.4 applied on sequence $\{\mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$ which fulfils conditions in FCLT B.10 (see note above Lemma 5.4) it holds

$$\sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \left(\sum_{t=1}^{\lfloor T\tau \rfloor} \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t - \frac{\lfloor T\tau \rfloor}{T} \sum_{t=1}^T \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \right) - \boldsymbol{\Sigma}_1 \mathbf{B}(\tau) \right\| = o_P(1),$$

for some $\boldsymbol{\Sigma}_1 > 0$, as $T \rightarrow \infty$. We now have to show that the error committed by replacing the parameters in the above formula with their maximum likelihood estimators is negligible:

$$\begin{aligned}
& \left(\sum_{t=1}^k \widehat{\mathbf{M}}_t^\top \boldsymbol{\Omega}^{-1} \widehat{\boldsymbol{\varepsilon}}_t - \sum_{t=1}^k \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \right) \\
& + \left(-\frac{k}{T} \sum_{t=1}^T \widehat{\mathbf{M}}_t^\top \boldsymbol{\Omega}^{-1} \widehat{\boldsymbol{\varepsilon}}_t + \frac{k}{T} \sum_{t=1}^T \mathbf{M}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \right) =: \mathbf{R}_{k,T} + \mathbf{S}_{k,T}.
\end{aligned}$$

In $\mathbf{R}_{k,T}$ and $\mathbf{S}_{k,T}$, the subscript k signalizes the summation boundary and T signalizes that all maximum likelihood estimators are based on the full sample $1, \dots, T$.

Let us analyze those terms:

$$\begin{aligned}
 \mathbf{R}_{k,T} &= \sum_{t=1}^k \left[\left((\mathbf{v}_t - \boldsymbol{\mu}_p - \widehat{\boldsymbol{\mu}}_p + \boldsymbol{\mu}_p) \otimes \mathbf{I}_n \right) \boldsymbol{\Omega}^{-1} (\widehat{\boldsymbol{\varepsilon}}_t - \boldsymbol{\varepsilon}_t + \boldsymbol{\varepsilon}_t) \right] - \\
 &\quad - \sum_{t=1}^k \left[\left((\mathbf{v}_t - \boldsymbol{\mu}_p) \otimes \mathbf{I}_n \right) \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \right] = \\
 &= \sum_{t=1}^k \left[\left((\mathbf{v}_t - \boldsymbol{\mu}_p) \otimes \mathbf{I}_n \right) \boldsymbol{\Omega}^{-1} (\widehat{\boldsymbol{\varepsilon}}_t - \boldsymbol{\varepsilon}_t) \right] + \sum_{t=1}^k \left[\left((\boldsymbol{\mu}_p - \widehat{\boldsymbol{\mu}}_p) \otimes \mathbf{I}_n \right) \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t \right] + \\
 &\quad + \sum_{t=1}^k \left[\left((\boldsymbol{\mu}_p - \widehat{\boldsymbol{\mu}}_p) \otimes \mathbf{I}_n \right) \boldsymbol{\Omega}^{-1} (\widehat{\boldsymbol{\varepsilon}}_t - \boldsymbol{\varepsilon}_t) \right] = \mathbf{R}_{k,T}^{(1)} + \mathbf{R}_{k,T}^{(2)} + \mathbf{R}_{k,T}^{(3)}.
 \end{aligned}$$

By Lemma 5.5 and the law of the iterated logarithm applied to iid sequence $\{\boldsymbol{\varepsilon}_t\}_{t \in \mathbb{Z}}$

$$\mathbf{R}_{k,T}^{(2)} = \left((\boldsymbol{\mu}_p - \widehat{\boldsymbol{\mu}}_p) \otimes \mathbf{I}_n \right) \boldsymbol{\Omega}^{-1} \sum_{t=1}^k \boldsymbol{\varepsilon}_t = \mathcal{O}(\ln \ln T) \quad \text{a.s., } T \rightarrow \infty.$$

Using Lemma 5.5 and arguments in the proof of Lemma 5.6 we get $\mathbf{R}_{k,T}^{(3)} = \mathcal{O}(\ln \ln T)$ a.s., $T \rightarrow \infty$. Utilizing expansion (5.12),

$$\mathbf{R}_{k,T}^{(1)} = - \sum_{t=1}^k \left((\mathbf{v}_t - \boldsymbol{\mu}_p) \otimes \mathbf{I}_n \right) \boldsymbol{\Omega}^{-1} (\mathbf{I}_n - \mathbf{K}) (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) - \quad (5.19)$$

$$- \sum_{j=1}^p \sum_{t=1}^k \left((\mathbf{v}_t - \boldsymbol{\mu}_p) \otimes \mathbf{I}_n \right) \boldsymbol{\Omega}^{-1} (\widehat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) (\mathbf{y}_{t-j} - \boldsymbol{\mu}) + \quad (5.20)$$

$$+ \sum_{j=1}^p \sum_{t=1}^k \left((\mathbf{v}_t - \boldsymbol{\mu}_p) \otimes \mathbf{I}_n \right) \boldsymbol{\Omega}^{-1} (\widehat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}). \quad (5.21)$$

It is immediate that term (5.19) is $\mathcal{O}(\ln \ln T)$ a.s., $T \rightarrow \infty$, and term (5.21) is $\mathcal{O}(T^{-\frac{1}{2}} (\ln \ln T)^{\frac{3}{2}})$ a.s., as $T \rightarrow \infty$. Term (5.20) is $n^2 p$ -vector of the form

$$\sum_{t=1}^k \sum_{j=1}^p \left[(\mathbf{y}_{t-j}^\top - \boldsymbol{\mu}^\top) \otimes (\mathbf{v}_t - \boldsymbol{\mu}_p) \otimes \mathbf{I}_n \right] \cdot \text{vec} \left(\boldsymbol{\Omega}^{-1} (\widehat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) \right).$$

Let

$$\boldsymbol{\rho} := \sum_{j=1}^p \left\{ \mathbb{E} \left[(\mathbf{y}_{t-j}^\top - \boldsymbol{\mu}^\top) \otimes (\mathbf{v}_t - \boldsymbol{\mu}_p) \otimes \mathbf{I}_n \right] \cdot \text{vec} \left(\boldsymbol{\Omega}^{-1} (\widehat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j) \right) \right\}.$$

Using FCLT B.10, it holds that term (5.20) is $k \cdot \boldsymbol{\rho} + \mathcal{O}(\ln \ln T)$ a.s., $T \rightarrow \infty$.

The $\mathbf{S}_{k,T}$ term can be analyzed in the similar way, because $\mathbf{S}_{k,T} = -\frac{k}{T} \mathbf{R}_{k,T}$. By adding $\mathbf{R}_{k,T}$ and $\mathbf{S}_{k,T}$, term $k\boldsymbol{\rho}$ (which is of order worse than $\mathcal{O}(\ln \ln T)$ a.s.) vanishes, hence we obtain $\sup_{1 \leq k \leq T} \|\mathbf{R}_{k,T} + \mathbf{S}_{k,T}\| = \mathcal{O}(\ln \ln T)$ a.s., $T \rightarrow \infty$, which concludes the proof of part (ii).

(iii) For the change detection in the variance structure of the VAR(p) model let us consider

$$\begin{aligned}
 \left[\frac{\partial}{\partial \Omega} \ell_k(\Omega) \right]_{\theta=\hat{\theta}} &= -\frac{k}{2} \hat{\Omega}^{-1} + \frac{1}{2} \hat{\Omega}^{-1} \left(\sum_{t=1}^k \hat{\varepsilon}_t \hat{\varepsilon}_t^\top \right) \hat{\Omega}^{-1} = \\
 &= -\frac{k}{2} \hat{\Omega}^{-1} \hat{\Omega} \hat{\Omega}^{-1} + \frac{1}{2} \hat{\Omega}^{-1} \left(\sum_{t=1}^k \hat{\varepsilon}_t \hat{\varepsilon}_t^\top \right) \hat{\Omega}^{-1} = \\
 &= \frac{1}{2} \hat{\Omega}^{-1} \left(-\frac{k}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t^\top + \sum_{t=1}^k \hat{\varepsilon}_t \hat{\varepsilon}_t^\top \right) \hat{\Omega}^{-1} = \\
 &= \frac{1}{2} \left[\Omega^{-1} \left(\sum_{t=1}^k \varepsilon_t \varepsilon_t^\top - \frac{k}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t^\top \right) \Omega^{-1} \right] + \\
 &+ \frac{1}{2} \left[\hat{\Omega}^{-1} \left(\sum_{t=1}^k \hat{\varepsilon}_t \hat{\varepsilon}_t^\top \right) \hat{\Omega}^{-1} - \Omega^{-1} \left(\sum_{t=1}^k \varepsilon_t \varepsilon_t^\top \right) \Omega^{-1} \right] - \\
 &- \frac{1}{2} \frac{k}{T} \left[\hat{\Omega}^{-1} \left(\sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t^\top \right) \hat{\Omega}^{-1} - \Omega^{-1} \left(\sum_{t=1}^T \varepsilon_t \varepsilon_t^\top \right) \Omega^{-1} \right] =: \\
 &=: \mathbf{A}_{k,T}^{(1)} + \mathbf{A}_{k,T}^{(2)} + \mathbf{A}_{k,T}^{(3)}.
 \end{aligned}$$

It holds that $\mathbf{A}_{k,T}^{(3)} = -\frac{k}{T} \mathbf{A}_{T,T}^{(2)}$. For $\mathbf{A}_{k,T}^{(1)}$, the FCLT yields

$$\sup_{0 \leq \tau \leq 1} \left\| \frac{1}{\sqrt{T}} \mathbf{A}_{[T\tau],T} - \Sigma_2 \mathbf{B}(\tau) \right\| = o_{\mathbb{P}}(1), \quad T \rightarrow \infty,$$

for some $\Sigma_2 > 0$. Since, as $T \rightarrow \infty$,

$$\mathbf{A}_{k,T}^{(2)} = \frac{1}{2} \Omega^{-1} \left[\sum_{t=1}^k \hat{\varepsilon}_t \hat{\varepsilon}_t^\top - \sum_{t=1}^k \varepsilon_t \varepsilon_t^\top \right] \Omega^{-1} + o(1) \quad \text{a.s.}, \quad (5.22)$$

due to Lemma 5.5, we can replace $\hat{\Omega}^{-1}$ with Ω^{-1} in $\mathbf{A}_{k,T}^{(2)}$ and also in $\mathbf{A}_{k,T}^{(3)}$.

Denoting $\mathbf{m} := (\mathbf{I}_n - \mathbf{K})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) - \sum_{j=1}^p (\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ and $\mathbf{r}_t := \sum_{j=1}^p (\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j)(\mathbf{y}_{t-j} - \boldsymbol{\mu})$, then $\hat{\varepsilon}_t = \varepsilon_t - \mathbf{r}_t - \mathbf{m}$. Hence term inside the square brackets of (5.22) can be expressed as

$$\begin{aligned}
 \sum_{t=1}^k (\hat{\varepsilon}_t \hat{\varepsilon}_t^\top - \varepsilon_t \varepsilon_t^\top) &= \sum_{t=1}^k \left(-\varepsilon_t \mathbf{r}_t^\top - \varepsilon_t \mathbf{m}^\top - \mathbf{r}_t \varepsilon_t^\top + \mathbf{r}_t \mathbf{r}_t^\top + \mathbf{r}_t \mathbf{m}^\top - \mathbf{m} \varepsilon_t^\top + \right. \\
 &\quad \left. + \mathbf{m} \mathbf{r}_t^\top + \mathbf{m} \mathbf{m}^\top \right).
 \end{aligned}$$

Applying “vec” operator on the above 8 terms, using Theorem 5.3, Law of the iterated logarithm for iid random vectors, and Lemma 5.5 we get, as $T \rightarrow \infty$,

$$\begin{aligned}
 \sum_{t=1}^k \text{vec}(\mathbf{r}_t \varepsilon_t^\top) &= \mathcal{O}(\ln \ln T) \quad \text{a.s.} \\
 \sum_{t=1}^k \text{vec}(\mathbf{m} \varepsilon_t^\top) &= \mathcal{O}(\ln \ln T) \quad \text{a.s.} \\
 \sum_{t=1}^k \text{vec}(\mathbf{m} \mathbf{m}^\top) &= \mathcal{O}(\ln \ln T) \quad \text{a.s.} \\
 \sum_{t=1}^k \text{vec}(\mathbf{r}_t \mathbf{m}^\top) &= \mathcal{O}\left(\frac{(\ln \ln T)^{\frac{3}{2}}}{T^{\frac{1}{2}}}\right) \quad \text{a.s.}
 \end{aligned}$$

As regards term $\text{vec}(\mathbf{r}_t \mathbf{r}_t^\top)$, we will proceed a little slower:

$$\begin{aligned}
 \sum_{t=1}^k \text{vec}(\mathbf{r}_t \mathbf{r}_t^\top) &= \sum_{t=1}^k \text{vec} \left(\sum_{j=1}^p \sum_{l=1}^p (\hat{\Phi}_j - \Phi_j)(\mathbf{y}_{t-j} - \boldsymbol{\mu})(\mathbf{y}_{t-l} - \boldsymbol{\mu})^\top \cdot \right. \\
 &\quad \left. \cdot (\hat{\Phi}_l - \Phi_l)^\top \right) = \\
 &= \sum_{j=1}^p \sum_{l=1}^p \left\{ \left[(\hat{\Phi}_l - \Phi_l) \otimes (\hat{\Phi}_j - \Phi_j) \right] \cdot \sum_{t=1}^k \left(\text{vec} \left((\mathbf{y}_{t-j} - \boldsymbol{\mu})(\mathbf{y}_{t-l} - \boldsymbol{\mu})^\top \right) - \right. \right. \\
 &\quad \left. \left. - \mathbb{E} \left[\text{vec} \left((\mathbf{y}_{t-j} - \boldsymbol{\mu})(\mathbf{y}_{t-l} - \boldsymbol{\mu})^\top \right) \right] \right) \right\} + \\
 &\quad + k \sum_{j=1}^p \sum_{l=1}^p \left[(\hat{\Phi}_l - \Phi_l) \otimes (\hat{\Phi}_j - \Phi_j) \right] \cdot \mathbb{E} \left[\text{vec} \left((\mathbf{y}_{t-j} - \boldsymbol{\mu})(\mathbf{y}_{t-l} - \boldsymbol{\mu})^\top \right) \right] = \\
 &= k \sum_{j=1}^p \sum_{l=1}^p \left[(\hat{\Phi}_l - \Phi_l) \otimes (\hat{\Phi}_j - \Phi_j) \right] \cdot \mathbb{E} \left[\text{vec} \left((\mathbf{y}_{t-j} - \boldsymbol{\mu})(\mathbf{y}_{t-l} - \boldsymbol{\mu})^\top \right) \right] \quad (5.23) \\
 &\quad + \mathcal{O} \left(\frac{(\ln \ln T)^{\frac{3}{2}}}{T^{\frac{1}{2}}} \right) \quad \text{a.s.}, \quad T \rightarrow \infty.
 \end{aligned}$$

Since term (5.23) will cancel with the same term when expanding $\mathbf{A}_{k,T}^{(3)}$ and the rest of terms are $\mathcal{O}(\ln \ln T)$ a.s., $T \rightarrow \infty$, then $\sup_{1 \leq k \leq T} \|\mathbf{A}_{k,T}^{(2)} + \mathbf{A}_{k,T}^{(3)}\| = \mathcal{O}(\ln \ln T)$ a.s., $T \rightarrow \infty$, and hence the theorem is proven. \square

5.6 Simulation study

As in the previous chapters we present some computational details which document the quality of the convergence of the score test statistic under H_0 . The simulation concept remains the same as in the previous chapters. We will concentrate on 2-dimensional VAR(1) model with matrices $\Phi^{(1)}$ and $\Phi^{(2)}$, see Section 3.5 for the details about notation.

5.6.1 Simulations under the null hypothesis

Let us begin with the situation when a change point is known apriori and simulate (for different $\tau = \frac{1}{8}, \frac{1}{2}, \frac{7}{8}$) test statistic $|\hat{B}_j(\tau)|$, $j = 1, \dots, r$ where $r = 9$, see the beginning of Section 5.2 for the definition of r ; $\hat{\mathbf{B}}(\tau)$ is defined in (5.3). Let us consider processes P_1 and P_2 of length $T = 200$ and errors [E1]–[E4]. The general concept is comparable to the situation in Subsection 3.5.1. Figure 5.2 depicts the empirical distribution function of $|\hat{B}_1(\tau)|$ (i.e. component of the test statistic belonging to μ_1) and distribution function of $|B_1(\tau)|$, where B_1 is univariate standard Brownian bridge. We can see very good performance for both processes P_1 and P_2 under errors [E1]–[E3] and worse results in case of [E4]. Under H_0 , practically the same can be seen across different values of τ . If we compare the results with the similar simulation study presented in Figure 3.1 the score test achieves better convergence for the more persistent autoregressive process P_2 . The simulation study shows the empirical results only for the first component $|\hat{B}_1(\tau)|$ of the score vector, however the performance of other components is almost the

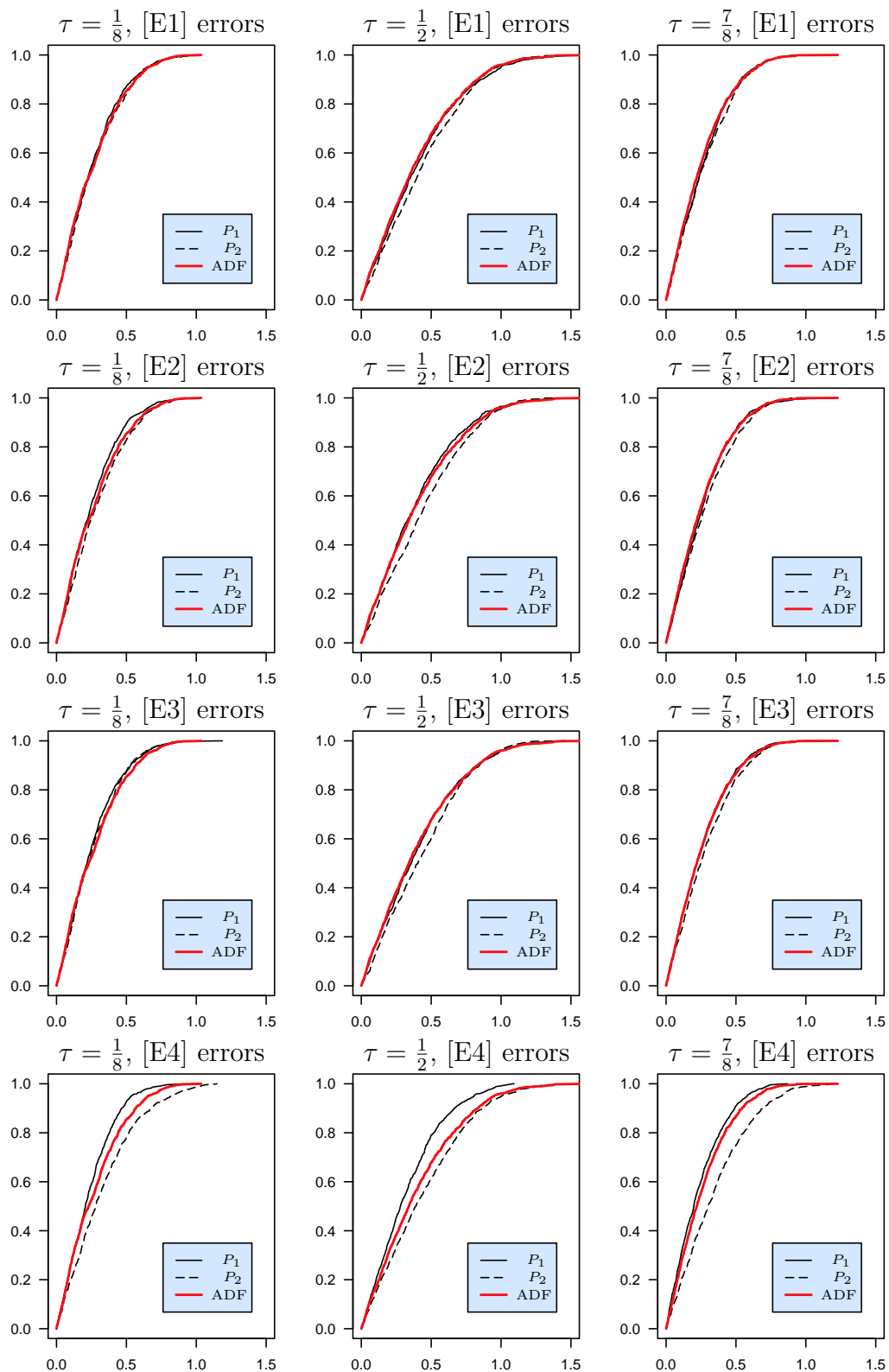


Figure 5.2: Situation when a change point is known: EDF of $|\widehat{B}_1(\tau)|$ (black lines) compared with ADF of $|B_1(\tau)|$ (red line) for $\tau = \frac{1}{8}, \frac{1}{2}, \frac{7}{8}$.

same. We do not show the figures for other components in order to save space. For particular break point τ with increasing length T , the convergence to the Brownian bridge is quite rapid as will become clear in the next simulation study

where we concentrate on the case of unknown break point.

Let us consider more practical situation of unknown break date. Figures 5.3 and 5.4 show the quality of convergence under H_0 , where each black line on a particular figure represents the performance of a single component $\sup_{0 \leq \tau \leq 1} |\widehat{B}_j(\tau)|$, $j = 1, \dots, 9$ with respect to the distribution of $\sup_{0 \leq \tau \leq 1} |B(\tau)|$, with B being standard univariate Brownian bridge.

The convergence results are pretty same in all cases across different errors and different type of processes. Clearly the longer series the better convergence to the asymptotic distribution which can be seen by comparing the figures from the left to the right.

Finally we present here the simulation results when testing all the components of score vector jointly, hence simulations of

$$\max_{j=1, \dots, 9} \sup_{0 \leq \tau \leq 1} |\widehat{B}_j(\tau)|$$

which are shown in Figure 5.5. The results are practically the same as on previous Figure 5.3 showing a rapid convergence.

5.6.2 Simulations under various alternatives

Since there is again no theoretical result of the behaviour of the proposed testing scheme under alternative hypothesis we will try to fill this gap at least by showing some empirical evidence.

We start with the simulations of process P_1 of length $T = 1000$ with errors being iid [E1] and heteroscedastic [E3]. Let change point occur in the middle of the series at $k = \frac{T}{2} = 500$. In order we could compare the simulations with results in Subsection 3.5.5, we will not detect change in variance for now, hence the reduced test statistic

$$\widehat{\mathbf{B}}_{[\text{red}]}(\tau) := \frac{1}{\sqrt{T}} \cdot \mathcal{J}_{[\text{red}]}^{-\frac{1}{2}}(\widehat{\boldsymbol{\theta}}) \cdot \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\mu}} \ell_{[T\tau]}(\widehat{\boldsymbol{\theta}}) \\ \frac{\partial}{\partial \boldsymbol{\phi}} \ell_{[T\tau]}(\widehat{\boldsymbol{\theta}}) \end{pmatrix}, \quad 0 \leq \tau \leq 1, \quad (5.24)$$

will be used, where

$$\mathcal{J}_{[\text{red}]}(\widehat{\boldsymbol{\theta}}) := \begin{pmatrix} (\mathbf{I}_n - \widehat{\mathbf{K}})^\top \widehat{\boldsymbol{\Omega}}^{-1} (\mathbf{I}_n - \widehat{\mathbf{K}}) & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{\Gamma}}_y(0) \otimes \widehat{\boldsymbol{\Omega}}^{-1} \end{pmatrix}.$$

First, we will discuss change in \mathbf{c} from $(0, 0)^\top$ to $\tilde{\mathbf{c}} = (\tilde{c}_1, \tilde{c}_2)^\top$ with other parameters being unchanged. As in Subsection 3.5.5 the empirical power under [E1] and [E3] will be compared. An empirical test will detect whether

$$\max_{j=1, \dots, 6} \sup_{0 \leq \tau \leq 1} |\widehat{B}_{[\text{red}],j}(\tau)| > 1.652,$$

where 1.652 is the critical value from Table 5.2 for testing jointly $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}^{(1)}$ on level $\alpha = 0.05$.

Graphically in Figure 5.6 we can see the power function under the alternative hypothesis when the parameter \mathbf{c} changes from the origin to the different vector. On x axis there is the Euclidean distance of new vector $\tilde{\mathbf{c}}$ from origin.

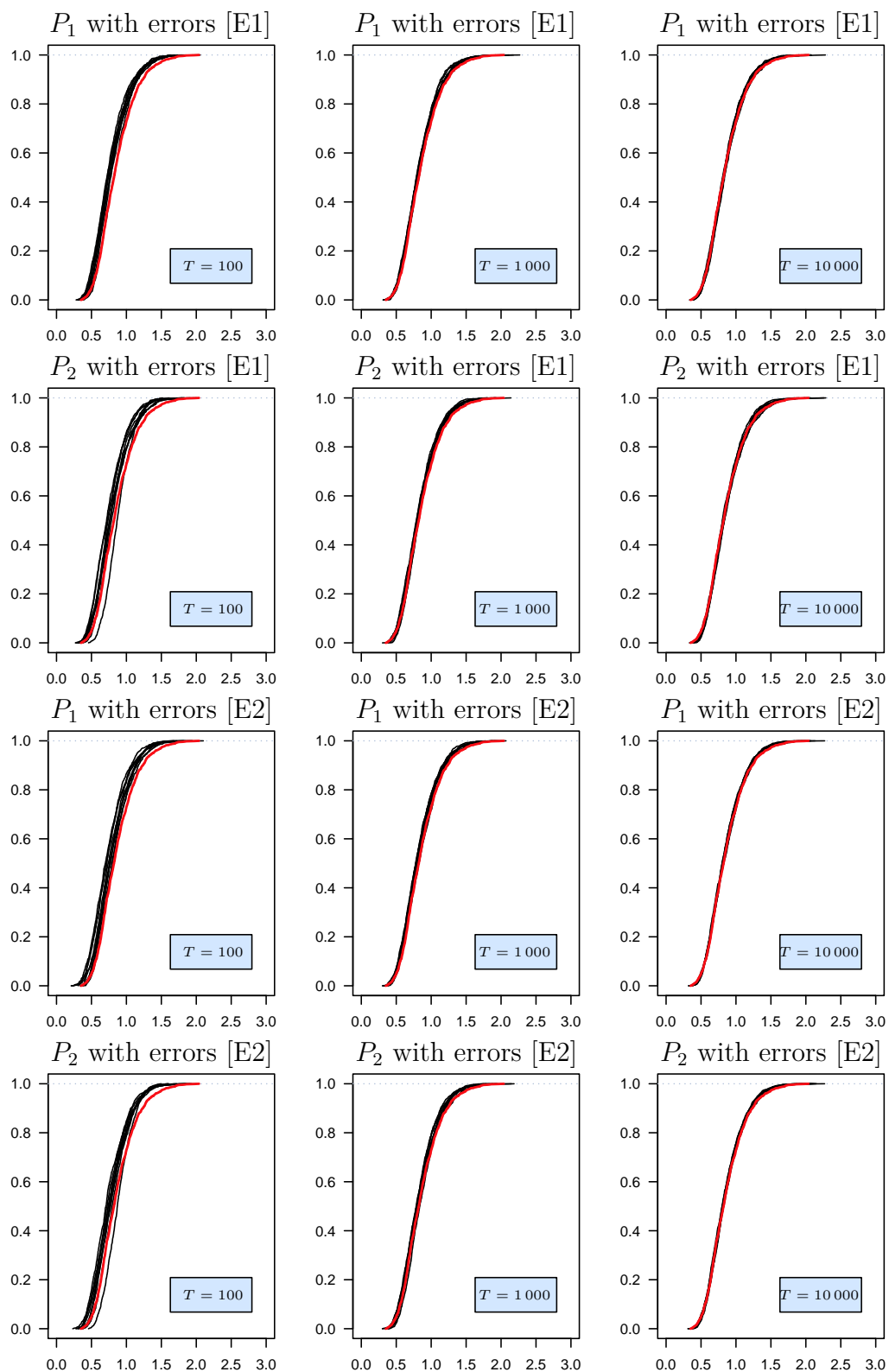


Figure 5.3: EDF for each $\sup_{0 \leq \tau \leq 1} |\hat{B}_j(\tau)|$, $j = 1, \dots, 9$, (black lines) compared to $\sup_{0 \leq \tau \leq 1} |B(\tau)|$ (red line), for processes P_1 and P_2 . Different T and errors [E1] and [E2] considered.

We can see that the heteroscedasticity in the error term causes a substantial decrease in terms of power of the test for changes in intercept. Approximately, test

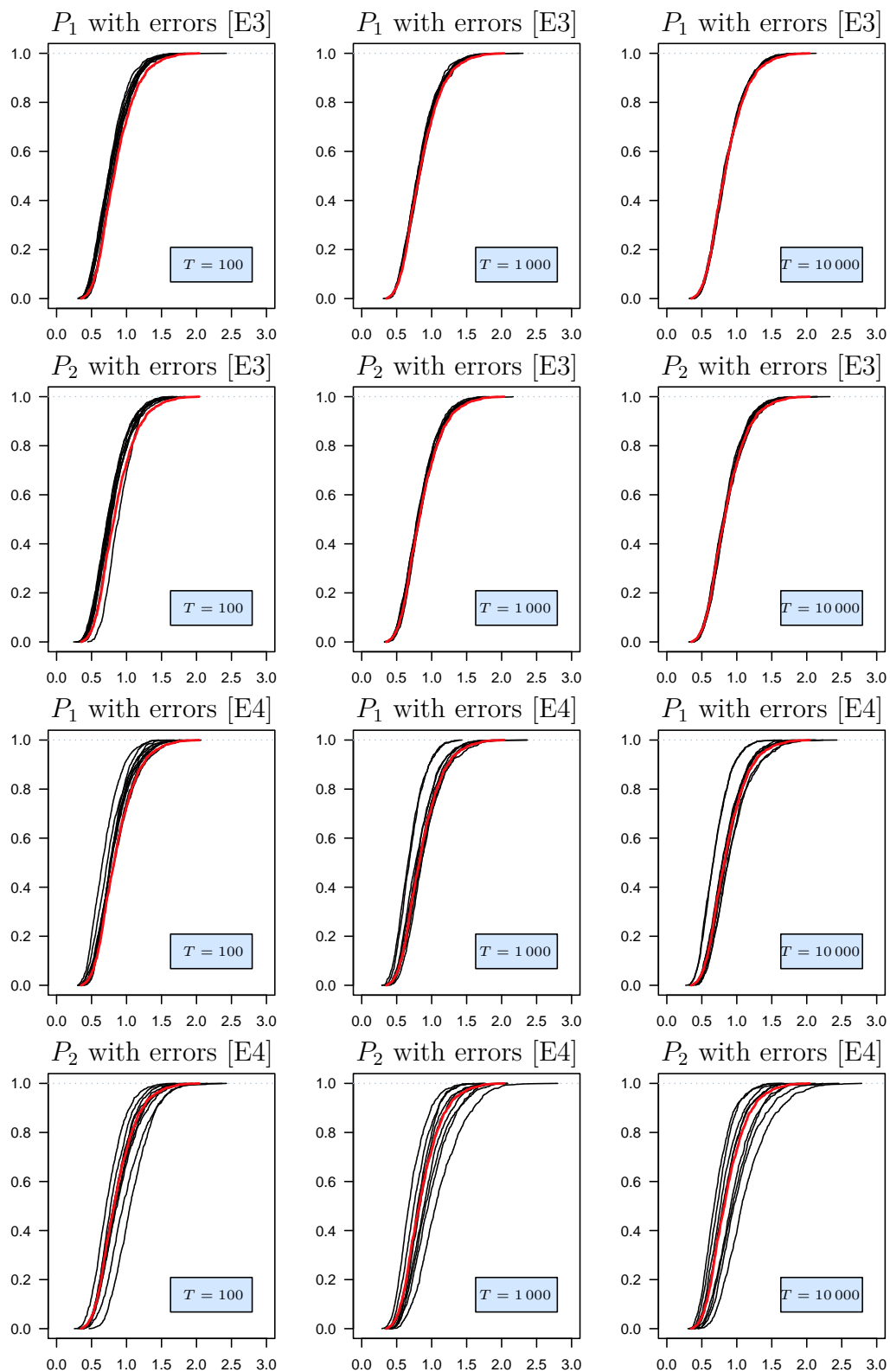


Figure 5.4: EDF for each $\sup_{0 \leq \tau \leq 1} |\widehat{B}_j(\tau)|$, $j = 1, \dots, 9$, (black lines) compared to $\sup_{0 \leq \tau \leq 1} |B(\tau)|$ (red line), for processes P_1 and P_2 . Different T and errors [E3] and [E4] considered.

reaches empirical power one for three as big change under heteroscedasticity than under iid errors. The biggest decrease of power as well as the lowest distance for

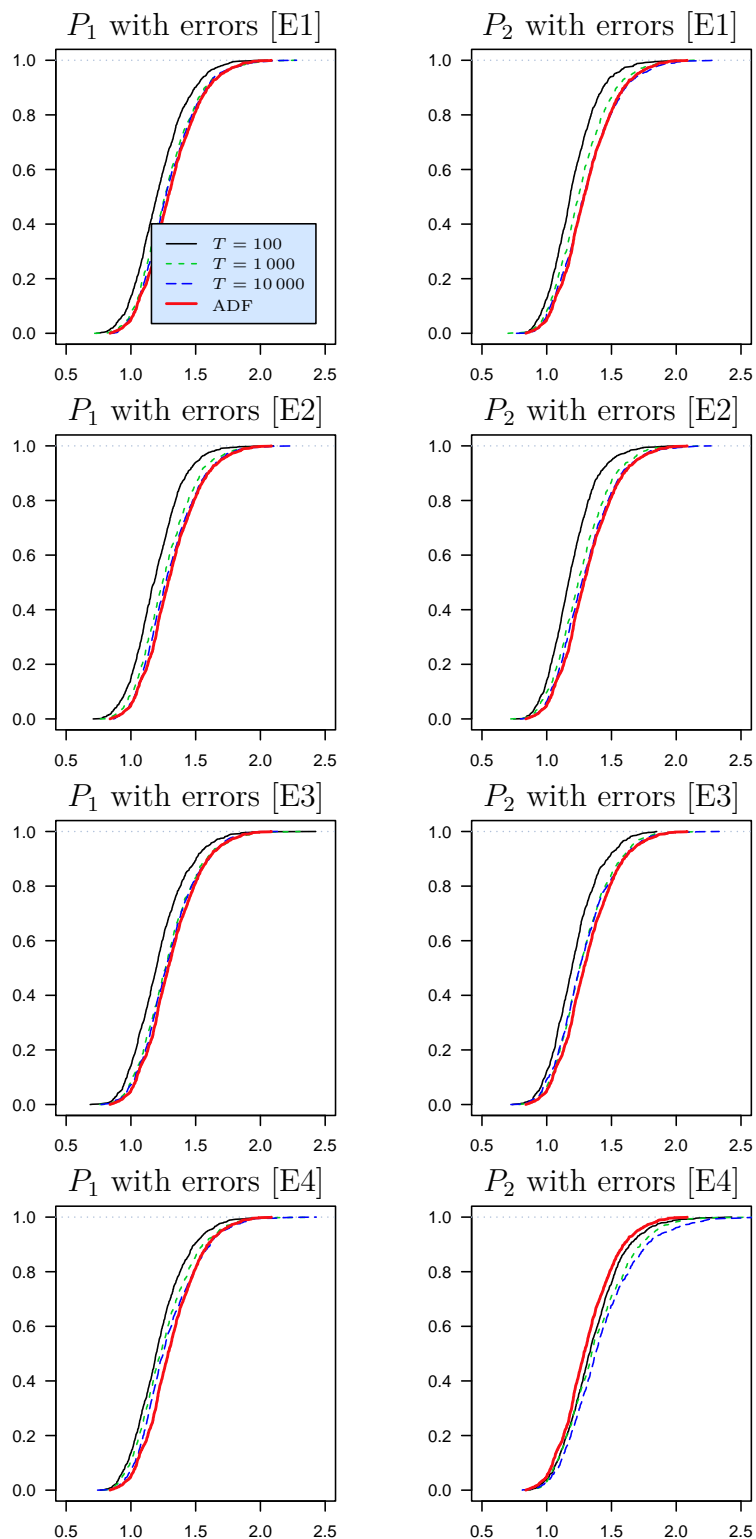
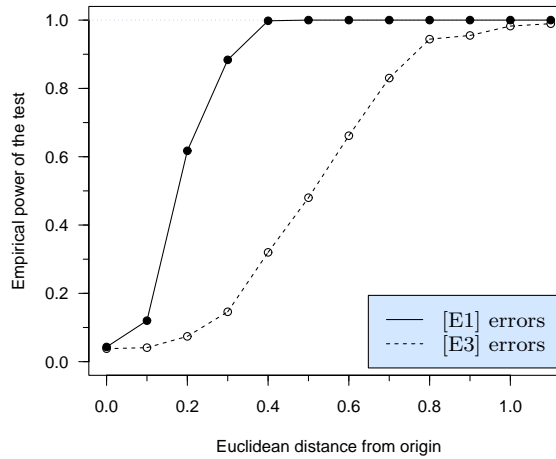


Figure 5.5: EDF for $\max_{1,\dots,9} \sup_{0 \leq \tau \leq 1} |\widehat{B}_j(\tau)|$, for different T , compared to $\max_{1,\dots,9} \sup_{0 \leq \tau \leq 1} |B(\tau)|$ (red line), for processes P_1 and P_2 and errors [E1]-[E4].

which the test firstly have empirical power one are highlighted in blue in Table 5.6. The score test is therefore more sensitive on the heteroscedasticity compared to Table 3.3 or Figure 3.6. The comparison with the latter results also reveals that in case of [E1] the score test reaches power one for somewhat smaller Euclidean

Table 5.3: Empirical sizes of the test.

\tilde{c}_1	\tilde{c}_2	$\sqrt{\tilde{c}_1^2 + \tilde{c}_2^2}$	power [E1]	power [E3]	% decr.
0.06	-0.05	0.08	0.10	0.04	-60
-0.05	0.16	0.17	0.43	0.07	-84
-0.16	0.12	0.20	0.55	0.07	-87
0.20	0.18	0.27	0.66	0.15	-77
0.03	0.36	0.37	1.00	0.26	-74
-0.42	0.05	0.42	1.00	0.33	-67
0.38	-0.26	0.46	1.00	0.40	-60
0.45	-0.17	0.48	1.00	0.52	-48
0.46	0.44	0.64	1.00	0.67	-33
0.96	0.76	1.23	1.00	1.00	0

Figure 5.6: Empirical power simulated for process $[P_1]$ of length $T = 1000$ with $[E1]$ and $[E3]$ error term sequences.

distance from origin compared to the likelihood ratio test, however approximately same Euclidean distance is needed for approaching power 1 under $[E3]$ for both tests.

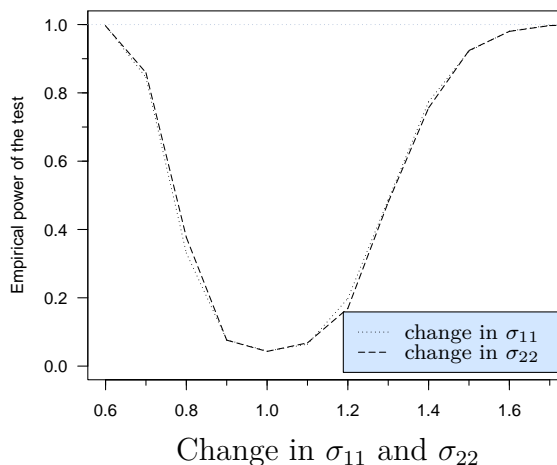
Empirical sizes of the test when only autoregression parameters change are depicted in Table 5.4 which can be compared with Table 3.4. The difference between the value of element of the autoregression matrix after and before the change is marked in blue scale. The darker the color the bigger the change.

We can again see that heteroscedasticity did not influence the empirical power too much which is the same conclusion as in Table 3.4. Comparing the power with results in that Subsection one might see higher power of the test both under $[E1]$ and under $[E3]$.

Finally we add the simulation results in case where components of variance matrix are changing. Figure 5.7 shows the empirical power under the changes in diagonal components of Ω .

Table 5.4: Empirical sizes of the test.

$\tilde{\varphi}_{11}$	$\tilde{\varphi}_{12}$	$\tilde{\varphi}_{21}$	$\tilde{\varphi}_{22}$	power [E1]	power [E3]
0.5	0.2	0.2	0.1	0.04	0.04
0.5	0.2	0.2	0.0	0.17	0.15
0.5	0.2	0.2	0.2	0.13	0.13
0.5	0.0	0.2	0.1	0.69	0.61
0.5	0.2	0.2	0.3	0.70	0.67
0.3	0.2	0.2	0.1	0.75	0.72
0.5	0.2	0.0	0.1	0.82	0.79
0.5	0.2	0.2	0.4	0.98	0.98
0.2	0.2	0.2	0.1	0.99	0.99
0.5	0.2	0.2	0.5	1.00	1.00
0.4	0.2	0.2	0.0	0.25	0.25
0.5	0.1	0.0	0.1	0.81	0.81
0.3	0.2	0.2	0.3	0.93	0.90
0.5	0.4	0.4	0.1	0.94	0.94
0.2	0.2	0.2	0.0	0.99	0.98
0.5	0.3	0.5	0.1	1.00	1.00

Figure 5.7: Empirical power simulated for process $[P_1]$ of length $T = 1000$ with $[E1]$ error term sequence.

5.7 Chapter summary

We presented a new score type test for change detection in stationary vector autoregressions. The idea of Gombay (2008) was extended to cover multivariate autoregressions with no restrictions imposed on the distribution of the error term by introducing a better standardization matrix. The test is very versatile since it enables to test either the parameters separately or in any combination depending on the needs of analysts. Simulation results show that the testing procedure achieves better results than the likelihood ratio test both under the null hypothesis and the alternatives for considered scenarios.

6. Conclusion

At the end, we would like to recapitulate a little bit and summarize the main contributions of this thesis: Chapter 2 brought together various sets of conditions which has appeared rather separately in the recent articles dealing with change-points. The aim was to point out at least to some implications of these assumptions in one place before tackling the main issues of this thesis and to provide some comparisons. Even if all the conditions can be found in the different sources we brought several new proofs. We should point out at least the proof of speed of decay of the sum of the norms of matrices of linear process which comes from the stationary vector autoregression. Another valuable contribution was the transmission of conditions (C.2) and (C.3) concerning error term onto the time series itself under conditions C which yields FCLT for such time series.

The last mentioned result was especially useful for Chapter 3 where we discussed the change point detection based on the likelihood ratio under no change in variance of the error term. Even if the asymptotic result is widely known for a given set of assumptions, the main aim and effort of that chapter were the performance of the proofs under various sets of conditions. When reading the proofs of the main theorems, one can get the impression about the strength of the various conditions. Compared to Qu and Perron (2007), some of the conditions can be weakened since we had a more concrete specification for the model and hence the FCLT as assumptions could be avoided. A new result, presented in Dvořák and Prášková (2013) as well, is Darling-Erdős type test which had not been published elsewhere before.

Chapter 4 brought the Darling-Erdős test for the modification of the likelihood ratio statistic under SCENARIO 3 which was a new contribution. We also explained the reason why the statistic based on log-likelihood ratio does not follow asymptotically the Gumbel distribution. The result is more of theoretical importance since the convergence of the standardized test statistic under the null hypothesis is very slow even for larger datasets. Perhaps more useful results might be gained by using resampling techniques.

Chapter 5 referred to a score test for multivariate stationary autoregression. This was a generalization of the univariate case proposed by Gombay (2008). We omitted the Gaussianity of the error term and extended the usage of the test to the non-Gaussian errors. The simulations showed the very rapid convergence even for smaller sample sizes. Comparison to the tests proposed in Chapter 3 revealed that the score test gives better results both under the null hypothesis and also in terms of the empirical power.

We want to say a few comments about the simulation part as well. The aim of the simulations stood in the documentation of the asymptotic results under the null and alternative hypotheses. Even it might seem strange for the reader, a big contribution of the text was also a study of the performance and visualization of the tests for long autoregression with $T = 10\,000$ observations. Computers can usually handle much larger data files, however this is not the case of free software R in case of larger dimensional problems. On a standard laptop, it is impossible to get least squares estimates even for such a simple model as 2 dimensional artificial VAR(1) model of $T = 10\,000$ observations even when we tried to use packages,

extend the internal memory of the computer or to use our own programming code with QR decompositions. Since we could benefit from using the computational cluster we were able to bring results even for larger time series. However, since only the basic form of \mathbf{R} without any advanced packages is installed there, we had to program all the codes by ourselves.

We should mention also a few comments about further developments which can be done. Of the primary interest is mainly the consistency result, i.e. whether the tests detect a change under the alternative hypothesis with the asymptotic power approaching one. Similarly as in Hušková et al. (2007), the theoretical result can be achieved by appropriate splitting the test statistic to different parts based on the presence/absence of the shift vector $\boldsymbol{\delta} = \boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}$ and showing that, under alternative, the components containing $\boldsymbol{\delta}$ approaching to infinity in probability, whereas the other components remain at least asymptotically bounded in probability. Under SCENARIO 1 the part of the proof has partly been done and the statistic has been decomposed under the alternative where remaining and the most difficult task remained to show the asymptotic result. The consistency of the test statistics based on the weighted residuals were simpler for proving consistency, see Hušková et al. (2007) for details. Roughly speaking, with omitting further notation details, their statistic is of the form $\mathbf{s}_k^\top \mathbf{H} \mathbf{s}_k$, \mathbf{s}_k are the weighted residuals, whereas in our case the quasi-likelihood statistic contains 3 addends $\mathbf{s}_k^\top \mathbf{H}_k \mathbf{s}_k + \tilde{\mathbf{s}}_k^\top \tilde{\mathbf{H}}_k \tilde{\mathbf{s}}_k - \mathbf{s}_T^\top \mathbf{H}_T \mathbf{s}_T$, and with different signs. In case of former statistic authors split the product to suitable addends and show that all but one addend are bounded in probability and the one addend tends to infinity. However, following this idea in the quasi-likelihood approach leads to the complicated structure of the test, since there are more “candidate” addends for approaching infinity. In addition, they have opposite signs and hence the conclusion about the consistency is more complicated. The desired “good” behaviour of the tests under various alternative hypotheses were documented in the simulation studies.

Another direction for further research can be weakening the independence of the error term in Chapter 5. This might be done via the replacement of independence by the weak dependence structures with some other moment restrictions. The key FCLTs and the rate of convergence of the estimators will remain valid under Assumptions A, B as well.

A. Matrix algebra

Here we remind some basic calculations with matrices, especially formulas for the Kronecker product, traces, vec of the matrices.

Lemma A.1 *Let $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ be the real matrices of appropriate dimensions such that all operations are correctly specified. Then the following holds:*

$$(1) \text{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A}) \cdot \text{vec}(\mathbf{B})$$

$$(2) (\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A}^\top \otimes \mathbf{B}^\top$$

$$(3) (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$$

$$(4) \text{If } \mathbf{A}, \mathbf{B} \text{ are invertible then } (\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$$

$$(5) (\mathbf{A} + \mathbf{B}) \otimes (\mathbf{C} \otimes \mathbf{D}) = \mathbf{A} \otimes \mathbf{C} + \mathbf{A} \otimes \mathbf{D} + \mathbf{B} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{D}$$

$$(6) \text{If } \mathbf{x} \text{ is a vector and } \mathbf{A} \text{ a square matrix of appropriate dimensions then } \mathbf{x}^\top \mathbf{Ax} = \text{tr}\{\mathbf{Axx}^\top\} \text{ and } \mathbf{x}^\top \mathbf{Ax} = (\mathbf{x}^\top \otimes \mathbf{x}^\top) \cdot \text{vec}(\mathbf{A})$$

Proof: The assertions can be found in Appendix A.11 and A.12 of Lütkepohl (2005). \square

Lemma A.2 *Let $\mathbf{x} \in \mathbb{R}^d$ be a vector, and $\mathbf{A} \in \mathbb{R}^{d \times d}$ be a matrix. Then*

$$(\mathbf{x}^\top \mathbf{Ax})^2 = \text{tr} \left\{ [(\text{vec } \mathbf{A})(\text{vec } \mathbf{A})^\top] \cdot (\mathbf{xx}^\top \otimes \mathbf{xx}^\top) \right\}.$$

Proof: Following rule (6) in previous Lemma A.1 we have

$$\begin{aligned} (\mathbf{x}^\top \mathbf{Ax})^2 &= [(\mathbf{x}^\top \otimes \mathbf{x}^\top) \cdot \text{vec } \mathbf{A}] \cdot [(\mathbf{x}^\top \otimes \mathbf{x}^\top) \cdot \text{vec } \mathbf{A}]^\top = \\ &= (\mathbf{x}^\top \otimes \mathbf{x}^\top)(\text{vec } \mathbf{A})(\text{vec } \mathbf{A})^\top(\mathbf{x} \otimes \mathbf{x}) = \\ &= \text{tr} \left\{ (\mathbf{x}^\top \otimes \mathbf{x}^\top)(\text{vec } \mathbf{A})(\text{vec } \mathbf{A})^\top(\mathbf{x} \otimes \mathbf{x}) \right\} = \\ &= \text{tr} \left\{ ((\text{vec } \mathbf{A})(\text{vec } \mathbf{A})^\top)((\mathbf{x} \otimes \mathbf{x})(\mathbf{x}^\top \otimes \mathbf{x}^\top)) \right\} = \\ &= \text{tr} \left\{ ((\text{vec } \mathbf{A})(\text{vec } \mathbf{A})^\top)(\mathbf{xx}^\top \otimes \mathbf{xx}^\top) \right\}. \end{aligned}$$

\square

B. Probability structures and asymptotic results

We mention the definitions of the probability structures and related limit theorems that were used in the text. We begin with martingale-difference sequences.

B.1 Martingale-differences

Definition B.1 Let $\{\mathcal{F}_t\}_{-\infty}^{\infty}$ be a filtration on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. d -dimensional random sequence $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ such that \mathbf{X}_t is adapted to \mathcal{F}_t is called martingale-difference sequence with respect to \mathcal{F}_t , if for all $t \in \mathbb{Z}$

$$\begin{aligned} \mathbb{E} \|\mathbf{X}_t\| &< \infty, \\ \mathbb{E}[\mathbf{X}_t | \mathcal{F}_{t-1}] &= \mathbf{0} \quad \text{a.s.} \end{aligned}$$

If $\mathcal{F}_t = \sigma\{\mathbf{X}_s, -\infty < s \leq t\}$ then it is simply referred to as *martingale-difference sequence*. It follows from Definition B.1 that the martingale-difference sequence is a centered sequence. Important property of martingale-differences is their uncorrelatedness, see for example Davidson (1994), Corollary 15.4.

Here we present the Central Limit Theorem for martingale-differences:

Theorem B.2 Let $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$ be an d -dimensional vector martingale-difference sequence. Suppose that

- $\mathbb{E}[\mathbf{Y}_t \mathbf{Y}_t^\top] = \Omega_t$ is positive-definite, with $T^{-1} \sum_{t=1}^T \Omega_t \xrightarrow{T \rightarrow \infty} \Omega$, and Ω is positive-definite as well,
- $\mathbb{E}[Y_{it} Y_{jt} Y_{kt} Y_{\ell t}] < \infty, \forall t \in \mathbb{Z}$, and $\forall (i, j, k, \ell)$, where Y_{it} is the i -th element of the vector \mathbf{Y}_t , and
- $T^{-1} \sum_{t=1}^T \mathbf{Y}_t \mathbf{Y}_t^\top \xrightarrow{T \rightarrow \infty} \Omega$.

Then

$$\frac{1}{\sqrt{T}} \cdot \sum_{t=1}^T \mathbf{Y}_t \xrightarrow{T \rightarrow \infty} \mathcal{N}(\mathbf{0}, \Omega).$$

Proof: See Hamilton (1994), p. 194, Proposition 7.9. □

In case of Assumptions A we used the the invariance principle for martingale difference arrays:

Theorem B.3 Let $\{\mathbf{X}_{T,t}, \mathcal{F}_{T,t}\}$ be an d -dimensional martingale difference array with variance matrix array $\{\Sigma_{T,t}\}$, such that $\sum_{t=1}^T \Sigma_{T,t} = \mathbf{I}_d$. If

- (i) $\sum_{t=1}^T \mathbf{X}_{T,t} \mathbf{X}_{T,t}^\top \xrightarrow{T \rightarrow \infty} \mathbf{I}_d$, as $T \rightarrow \infty$,
- (ii) $\max_{t=1, \dots, T} \mathbf{X}_{T,t}^\top \mathbf{X}_{T,t} \xrightarrow{T \rightarrow \infty} 0$, as $T \rightarrow \infty$,

(iii) $\sum_{t=1}^{\lfloor T\tau \rfloor} \mathbf{X}_{T,t} \rightarrow \tau \mathbf{1}_d$, $\forall \tau \in [0, 1]$, as $T \rightarrow \infty$,

then for

$$\mathbf{Y}_T(\tau) := \sum_{t=1}^{\lfloor T\tau \rfloor} \mathbf{X}_{T,t} + (T\tau - \lfloor T\tau \rfloor) \mathbf{X}_{T, \lfloor T\tau \rfloor + 1}, \quad \tau \in [0, 1],$$

it holds

$$\mathbf{Y}_T(\cdot) \xrightarrow[T \rightarrow \infty]{D_d[0,1]} \mathbf{W}(\cdot),$$

where \mathbf{W} denotes d -dimensional standard Wiener process and $\xrightarrow[T \rightarrow \infty]{D_d[0,1]}$ denotes the convergence in Skorohod space.

Proof: See Davidson (1994), p. 454, Theorem 27.17. □

B.2 Stationarity and ergodicity

Definition B.4 Random process $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is said to be strictly-stationary if and only if $\forall k, \forall t_1, \dots, t_k \in \mathbb{Z}$ and $\forall h \in \mathbb{Z}$, the random vectors $(\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_k})$ and $(\mathbf{X}_{t_1+h}, \dots, \mathbf{X}_{t_k+h})$ have the same joint distribution.

Let us now define the ergodic sequences which together with their strict stationarity and integrability imply the strong law of large numbers, see Theorem B.7 below.

Definition B.5 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $T : \Omega \rightarrow \Omega$ a one-to-one function such that both T and T^{-1} are measurable. Let for all $E \in \mathcal{F}$: $\mathbb{P}[T^{-1}E] = \mathbb{P}[E]$, i.e. T is a measure-preserving transformation (m.p.t.). Event $E \in \mathcal{F}$ is called invariant under transformation T if and only if $E = T^{-1}E$. M.p.t. T is said to be ergodic if and only if for any invariant event E , we have $\mathbb{P}[E] = 0$ or $\mathbb{P}[E] = 1$.

Let T be now a shift-operator. Strictly-stationary process $\{\mathbf{X}_t(\omega)\}_{t \in \mathbb{Z}}$ is said to be ergodic if and only if $\mathbf{X}_t(\omega) = \mathbf{X}_1(T^{t-1}(\omega))$ for any $t \in \mathbb{Z}$, where T is m.p.t. and ergodic.

Advantage is that a time-invariant measurable function of strictly stationary and ergodic \mathbf{X}_t is also strictly stationary and ergodic:

Theorem B.6 If $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is strictly stationary and ergodic sequence and ϕ is measurable function not depending on t , then $\mathbf{Y}_t := \phi(\dots, \mathbf{X}_{t-1}, \mathbf{X}_t, \mathbf{X}_{t+1}, \dots)$ is also strictly stationary and ergodic.

Proof: See Billingsley (1995), p. 495, Theorem 36.4. □

Ergodic theorem gives us the strong law of large numbers for strictly stationary and ergodic sequences.

Theorem B.7 Let $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ be a strictly stationary, ergodic and integrable sequence on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{X}_t = \mathbb{E}[\mathbf{X}_1] \quad \text{a.s.}$$

Proof: See Davidson (1994), p. 200, Theorem 13.12. □

It can be shown that ergodicity property is the necessary and sufficient condition under which strictly stationary integrable process obeys the strong law of large numbers. As shown in Kakutani and Petersen (1981) the speed of convergence in Theorem B.7 is arbitrarily slow.

B.3 Strong mixings

The concept of the strong mixing sequences was introduced by Rosenblatt (1956) in order to prove the central limit theorem for dependent sequences. Strong mixing transformations are useful in the ergodic theory as well, since they imply ergodicity. We remind the basic definition, see for instance Davidson (1994), p. 209:

Definition B.8 *Let $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ be a sequence of random vectors on probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Define $\mathcal{F}_a^b := \sigma\{\mathbf{X}_t, a \leq t \leq b\}$ as a σ -field generated by the random vectors $\{\mathbf{X}_a, \dots, \mathbf{X}_b\}$, $a, b \in \mathbb{Z}$. Define*

$$\alpha(k) := \sup_{n \in \mathbb{Z}} \sup_{\{E \in \mathcal{F}_{-\infty}^n, F \in \mathcal{F}_{n+k}^\infty\}} |\mathbf{P}[E \cap F] - \mathbf{P}[E]\mathbf{P}[F]|.$$

If $\alpha(k) \rightarrow 0$, as $k \rightarrow \infty$, then $\{\mathbf{X}_t\}$ is called a strong-mixing process.

Note 1 (Special cases): If \mathbf{X}_t are independent random vectors then for all $k \in \mathbb{N}$: $\alpha(k) = 0$. In case that sequence $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is m -dependent, then it holds that $\alpha(k) = 0$ for all $k > m$. If $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is a strictly stationary process then we can omit “ $\sup_{n \in \mathbb{Z}}$ ” in the definition of $\alpha(k)$.

Note 2 (Speed of convergence): The speed of convergence of $\alpha(k)$ to zero is often important in Definition B.8. As we know from the ergodic theory, without any further assumptions, the convergence of strictly stationary and ergodic random sequence is arbitrarily slow in the strong law of large numbers, see Kakutani and Petersen (1981). Therefore it is often important to assume a certain *rate* of convergence of $\alpha(k)$ for having the rate in e.g. FCLT.

In that case a useful proposition for random variables is stated in Theorem 14.1. of Davidson (1994), p. 210, that any measurable function of finite sequence of strong mixing random variables is again a strong mixing with the same rate:

Theorem B.9 *If $\mathbf{Y}_t = g(\mathbf{X}_t, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-\tau})$, where τ is finite and g is a measurable function, then if \mathbf{X}_t is a strong mixing of size $\alpha(m) = \mathcal{O}(m^{-\phi})$, then \mathbf{Y}_t is a strong mixing with the same size.*

Proof: Analogously as in Theorem 14.1. of Davidson (1994). □

We now formulate FCLT for the strong-mixing random vectors.

Theorem B.10 *Let $\{\mathbf{X}_t\}_{t \geq 1}$ be a weak stationary sequence of \mathbb{R}^d -valued random vectors centered at expectations and having $(2 + \delta)$ -moments, $0 < \delta \leq 1$, uniformly bounded by 1. Suppose that $\{\mathbf{X}_t\}_{t \geq 1}$ satisfies the strong-mixing condition with $\alpha(k) = \mathcal{O}(k^{-(1+\epsilon)(1+2/\delta)})$, $\epsilon > 0$. Then the two series in $\gamma_{i,j} = \mathbb{E}[X_{i,1}X_{j,1}] + \sum_{s \geq 2} \mathbb{E}[X_{i,1}X_{j,s}] + \sum_{s \geq 2} \mathbb{E}[X_{i,s}X_{j,1}]$ converge absolutely. Denote $\mathbf{\Gamma} = \{\gamma_{i,j}\}_{i,j}$. Then we can redefine the sequence $\{\mathbf{X}_t\}_{t \geq 1}$ on a possibly wider probability space together with Wiener process $\mathbf{W}_{\mathbf{\Gamma}}(t)$ with variance matrix $\mathbf{\Gamma}$ such that*

$$\sum_{n \leq t} \mathbf{X}_n - \mathbf{W}_{\mathbf{\Gamma}}(t) = \mathcal{O}(t^{\frac{1}{2}-\lambda}) \quad \text{a.s.}$$

with some $\lambda > 0$ depending on ϵ , δ and d only.

Proof: The proof can be found in Kuelbs and Philipp (1980), Theorem 4. \square

B.4 M-dependent structures

Theorem B.11 *Assume that the d -dimensional random process $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$ is specified as*

$$\mathbf{Y}_t = \mathbf{f}(\boldsymbol{\nu}_t, \boldsymbol{\nu}_{t-1}, \dots), \quad t \in \mathbb{Z},$$

where $\mathbf{f}: \mathbb{R}^{n' \times \infty} \rightarrow \mathbb{R}^n$ is a measurable function and $\{\boldsymbol{\nu}_t\}_{t \in \mathbb{Z}}$ a sequence of independent, identically distributed random vectors with values in $\mathbb{R}^{n'}$.

Let $\mathbb{E}[\mathbf{Y}_t] = \mathbf{0}$ and $\mathbb{E} \|\mathbf{Y}_t\|^2 < \infty$. Suppose further that, for any $m \geq 1$, the m -dependent vectors $\mathbf{Y}_0^{(m)}$ can be defined such that

$$\sum_{m=1}^{\infty} \left(\mathbb{E} \|\mathbf{Y}_0 - \mathbf{Y}_0^{(m)}\|^2 \right)^{\frac{1}{2}} < \infty.$$

Then the series $\mathbf{\Gamma} := \sum_{t \in \mathbb{Z}} \text{cov}(\mathbf{Y}_0, \mathbf{Y}_t)$ converges (coordinatewise) absolutely and

$$\frac{1}{\sqrt{T}} \cdot \sum_{t=1}^{\lfloor T\tau \rfloor} \mathbf{Y}_t \xrightarrow{D_d[0,1]} \mathbf{W}_{\mathbf{\Gamma}}(\tau), \quad T \rightarrow \infty,$$

where $\mathbf{W}_{\mathbf{\Gamma}}$ is a d -dimensional Wiener process with variance matrix $\mathbf{\Gamma}$.

Proof: See proof in Aue et al. (2009), Theorem A.1. \square

Theorem B.12 *Suppose that the assumptions of Theorem B.11 hold true and that $\mathbb{E} \|\mathbf{Y}_t\|^4 < \infty$. If*

$$\sum_{m=1}^{\infty} \left(\mathbb{E} \|\mathbf{Y}_0 - \mathbf{Y}_0^{(m)}\|^4 \right)^{\frac{1}{4}} < \infty$$

is satisfied, then the series

$$\mathbf{\Sigma} := \sum_{t \in \mathbb{Z}} \text{cov}(\text{vech}(\mathbf{Y}_0 \mathbf{Y}_0^{\top}), \text{vech}(\mathbf{Y}_t \mathbf{Y}_t^{\top}))$$

converges (coordinatewise) absolutely and

$$\frac{1}{\sqrt{T}} \cdot \sum_{t=1}^{\lfloor T\tau \rfloor} \left(\text{vech} \left(\mathbf{Y}_t \mathbf{Y}_t^\top - \mathbb{E}[\mathbf{Y}_t \mathbf{Y}_t^\top] \right) \right)^{D_{d'}^{[0,1]}} \mathbf{W}_\Sigma(\tau), \quad T \rightarrow \infty,$$

where $d' = \frac{d(d+1)}{2}$.

Proof: See proof in Aue et al. (2009), Theorem A.2. □

List of Abbreviations

We have used the following notation in the thesis:

\mathbb{N}	Set of positive integers
\mathbb{N}_0	Set of non-negative integers
\mathbb{Z}	Set of integers
\mathbb{R}	Set of real numbers
\mathbb{C}	Set of complex numbers
$[\cdot]$	Integer part of a number
\ln	Natural logarithm
$\mathbb{I}_{[\text{cond}]}$	indicator function; $\mathbb{I}_{[\text{cond}]} = 1$ when condition "cond" is met, 0 otherwise
\mathbf{v}^\top	transposition of (column) vector \mathbf{v}
$\ \mathbf{v}\ $	Euclidean norm of a d -vector, i.e. $\ \mathbf{v}\ = \sqrt{\sum_{i=1}^d v_i^2}$
$\text{vec}(\mathbf{A})$	vector $(\mathbf{a}_1^\top, \dots, \mathbf{a}_d^\top)^\top$, where $\mathbf{a}_1, \dots, \mathbf{a}_d$ are the column vectors of matrix \mathbf{A}
$\text{vech}(\mathbf{A})$	(for squared symmetric matrices): vector of elements of matrix on and below the main diagonal, i.e. for $\mathbf{A} \in \mathbb{R}^{d \times d}$, $\text{vech}(\mathbf{A}) = (a_{11}, \dots, a_{d1}, a_{22}, \dots, a_{d2}, \dots, a_{dd})^\top$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of matrices \mathbf{A} and \mathbf{B}
$\det\{\mathbf{A}\}$	determinant of matrix \mathbf{A}
$\text{tr}\{\mathbf{A}\}$	trace of the matrix \mathbf{A} , i.e. sum of its diagonal elements
$\mathbf{A} > 0$	positive-definite matrix
$\lambda_{\min}(\mathbf{A})$	smallest eigenvalue of the matrix \mathbf{A}
$\lambda_{\max}(\mathbf{A})$	largest eigenvalue of the matrix \mathbf{A}
a.s.	almost surely; it means that the statement does not hold only on a set of probability measure 0
$\stackrel{d}{\equiv}$	equals in distribution
$\xrightarrow{d}_{T \rightarrow \infty}$	converge in distribution as $T \rightarrow \infty$
$\xrightarrow{P}_{T \rightarrow \infty}$	converge in probability as $T \rightarrow \infty$
$\xrightarrow{D_d[0,1]}_{T \rightarrow \infty}$	convergence in Skorohod space $D_d[0, 1]$, as $T \rightarrow \infty$, where d is dimension of the corresponding vector
$\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Omega})$	d -dimensional normal random variable with expected value $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Omega}$.
$\mathcal{O}(\cdot), o(\cdot)$	Landau symbols for almost sure boundedness and convergence
$\mathcal{O}_P(\cdot), o_P(\cdot)$	Landau symbols for boundedness and convergence in probability
ADF	asymptotic distribution function
ARMA	autoregressive moving average
CUSUM	cumulative sum
EDF	empirical distribution function
FCLT	Functional Central Limit Theorem
iid	independent identically distributed
std	standard deviation

References

- Andrews, D. W. (1993). Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, *61*(4), 821–856.
- Aue, A., Hörmann, S., Horváth, L., & Reimherr, M. (2009). Break Detection in the Covariance Structure of Multivariate Time Series Model. *The Annals of Statistics*, *37*(6B), 4046–4087.
- Aue, A., & Horváth, L. (2012). Structural Breaks in Time Series. *Journal of Time Series Analysis*, *34*(1), 1–16.
- Aue, A., Horváth, L., Hušková, M., & Kokoszka, P. (2006). Change Point Monitoring in Linear Models. *Econometrics Journal*, *9*, 373–403.
- Bagshaw, M., & Johnson, R. A. (1977). Sequential Procedures for Detecting Parameter Changes in a Time-Series Model. *Journal of the American Statistical Association*, *72*(359), 593–597.
- Bai, J. (1997). Estimation of a Change Point in Multiple Regression Models. *The Review of Economics and Statistics*, *79*(4), 551–563.
- Bai, J. (2000). Vector Autoregressive Models with Structural Changes in Regression Coefficients and in Variance-Covariance Matrices. *Annals of Economics and Finance*, *1*(2), 303–339.
- Bai, J., Lumsdaine, R. L., & Stock, J. H. (1998). Testing For and Dating Common Breaks in Multivariate Time Series. *Review of Economic Studies*, *65*(3), 395–432.
- Bai, J., & Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, *66*(1), 47–78.
- Basseville, M., & Nikiforov, I. (1993). *Detection of Abrupt Changes: Theory and Application*. New Jersey: Prentice-Hall.
- Berkes, I., Horváth, L., & Kokoszka, P. (2004). Testing for Parameter Constancy in GARCH(p,q) Models. *Statistics and Probability Letters*, *70*(4), 263–273.
- Billingsley, P. (1995). *Probability and Measure* (3rd ed.). New York: Wiley.
- Billingsley, P. (1999). *Convergence of Probability Measures* (2nd ed.). New York: Wiley.
- Chan, J., Horváth, L., & Hušková, M. (2013). Darling-Erdős Limit Result for Change-point Detection in Panel Data. *Journal of Statistical Planning and Inference*, *143*, 955–970.
- Chochola, O., Hušková, M., Prášková, Z., & Steinebach, J. G. (2013). Robust Monitoring of CAPM Portfolio Betas. *Journal of Multivariate Analysis*, *115*, 374–395.
- Csörgő, M., & Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Chichester: Wiley.
- Csörgő, M., & Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Budapest: Akadémiai Kiadó.
- Darling, D. A., & Erdős, P. (1956). A Limit Theorem for the Maximum of Normalized Sums of Independent Random Variables. *Duke Math. J.*, *23*, 143–155.
- Davidson, J. (1994). *Stochastic Limit Theory*. New York: Oxford University Press.
- Davis, R. A., Huang, D., & Yao, Y.-C. (1995). Testing for a Change in the

- Parameter Values and Order of an Autoregressive Model. *The Annals of Statistics*, 23(1), 282–304.
- Davison, A. (2003). *Statistical Models*. New York: Cambridge University Press.
- Dvořák, M. (2015). Darling-Erdős Type Test for Change Detection in Stationary VAR Models. *to appear in Communication in Statistics – Theory and Methods*.
- Dvořák, M., & Prášková, Z. (2013). On Testing Changes in Autoregressive Parameters of a VAR Model. *Communication in Statistics – Theory and Methods*, 42, 1208–1226.
- Fisher, R. A. (1925). Theory of Statistical Estimation. *Proceedings of the Cambridge Philosophical Society*, 22(5), 700–725.
- Gardner, L. A. (1969). On Detecting Changes in the Mean of Normal Variates. *The Annals of Mathematical Statistics*, 40(1), 116–126.
- Gombay, E. (2008). Change Detection in Autoregressive Time Series. *Journal of Multivariate Analysis*, 99(3), 451–464.
- Gombay, E., & Horváth, L. (2009). Sequential Tests and Change Detection in the Covariance Structure of Weakly Stationary Time Series. *Communications in Statistics - Theory and Methods*, 38, 2872–2883.
- Gombay, E., & Serban, D. (2009). Monitoring Parameter Change in ar(p) Time Series Models. *Journal of Multivariate Analysis*, 100, 715–725.
- Hall, P. (1979). On the Rate of Convergence of Normal Extremes. *Journal of Applied Probability*, 16(2), 433–439.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Hinkley, D. V. (1970). Inference About the Change-Point in a Sequence of Random Variables. *Biometrika*, 57(1), 1–17.
- Hörmann, S., & Kokoszka, P. (2010). Weakly Dependent Functional Data. *The Annals of Statistics*, 38(3), 1845–1884.
- Horváth, L. (1993a). Change in Autoregressive Processes. *Stochastic Processes and their Applications*, 44(2), 221–242.
- Horváth, L. (1993b). The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations. *The Annals of Statistics*, 21(2), 671–680.
- Horváth, L. (1995). Detecting Changes in Linear Regressions. *Statistics: A Journal of Theoretical and Applied Statistics*, 26(3), 189–208.
- Horváth, L. (1997). Detection of Changes in Linear Sequences. *Ann. Inst. Statist. Math.*, 49(2), 271–283.
- Horváth, L., Hušková, M., Kokoszka, P., & Steinebach, J. (2004). Monitoring Changes in Linear Models. *Journal of Statistical Planning and Inference*, 126(1), 225–251.
- Horváth, L., Hušková, M., & Serbinowska, M. (1997). Estimators for the Time of Change in Linear Models. *Statistics: A Journal of Theoretical and Applied Statistics*, 29(2), 109–130.
- Horváth, L., Kokoszka, P., & Steinebach, J. (2007). On Sequential Detection of Parameter Changes in Linear Regression. *Statistics and Probability Letters*, 77, 885–895.
- Hušková, M., & Chochola, O. (2010). Simple Sequential Procedures for Change in Distribution. In J. Antoch, M. Hušková, & P. K. Sen (Eds.), *Nonpara-*

- metrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková* (Vol. 7, pp. 95–104). Beachwood, Ohio, USA: Institute of Mathematical Statistics.
- Hušková, M., & Koubková, A. (2005). Monitoring Jump Changes in Linear Models. *Journal of Statistical Research*, 39(2), 51–70.
- Hušková, M., & Pícek, J. (2002). M-tests for Detection of Structural Changes in Regression. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L1-norm and Related Methods* (p. 213–227). Birkhäuser Basel.
- Hušková, M., Prášková, Z., & Steinebach, J. (2007). On the Detection of Changes in Autoregressive Time Series I. Asymptotics. *Journal of Statistical Planning and Inference*, 137(4), 1243–1259.
- Hušková, M., Prášková, Z., & Steinebach, J. (2008). On the Detection of Changes in Autoregressive Time Series II. Resampling procedures. *Journal of Statistical Planning and Inference*, 138(6), 1697–1721.
- James, B., James, K. L., & Siegmund, D. (1987). Tests for Change-points. *Biometrika*, 74(1), 71–84.
- Kakutani, S., & Petersen, K. (1981). The Speed of Convergence in the Ergodic Theorem. *Monatshefte für Mathematik*, 91(1), 11–18.
- Kokoszka, P., & Leipus, R. (2000). Change-point Estimation in ARCH Models. *Bernoulli*, 6(3), 513–539.
- Kotz, S., & Nadarajah, S. (2004). *Multivariate t Distribution and their Applications*. Cambridge: Cambridge University Press.
- Kuelbs, J., & Philipp, W. (1980). Almost Sure Invariance Principles for Partial Sums of Mixing B-Valued Random Variables. *The Annals of Probability*, 8(6), 1003–1036.
- Lee, S., Park, S., Maekawa, K., & Kawai, K. ichi. (2006). Test for Parameter Change in ARIMA Models. *Communications in Statistics - Simulation and Computation*, 35(2), 429–439.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.
- Nielsen, B. (2005). Strong Consistency Results for Least Squares Estimators in General Vector Autoregressions with Deterministic Terms. *Econometric Theory*, 21(3), 534–561.
- OOdaira, H., & Yoshihara, K. (1971). The Law of the Iterated Logarithm for Stationary Processes Satisfying Mixing Conditions. *Kodai Math. Sem. Rep.*, 23(3), 311–334.
- Page, E. (1955). A Test for a Change in a Parameter Occurring at an Unknown Point. *Biometrika*, 42(3/4), 523–527.
- Pham, T., & Tran, L. (1985). Some Mixing Properties of Time Series Models. *Stochastic Processes and Their Applications*, 19, 297–303.
- Picard, D. (1985). Testing and Estimating Change-Points in Time Series. *Advances in Applied Probability*, 17(4), 841–867.
- Prášková, Z. (2015). Monitoring Changes in RCA Models. *Springer Proceedings in Mathematics and Statistics*, 122, 129–137.
- Prášková, Z., & Chochola, O. (2014). M-procedures for Detection of a Change Under Weak Dependence. *Journal of Statistical Planning and Inference*, 149, 60–76.
- Qu, Z., & Perron, P. (2007). Estimating and Testing Structural Changes in

REFERENCES

- Multivariate Regressions. *Econometrica*, 75(2), 459–502.
- R Core Team. (2015). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Available from <http://www.R-project.org/>
- Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1), 43–47.
- Salazar, D. (1982). Structural Changes in Time Series Models. *Journal of Econometrics*, 19(1), 147–163.
- Tsay, R. (2010). *Analysis of Financial Time Series*. New Jersey: Wiley.
- Worsley, K. J. (1986). Confidence Regions and Tests for a Change-Point in a Sequence of Exponential Family Random Variables. *Biometrika*, 73(1), 91–104.
- Wu, W. B., & Min, W. (2005). On Linear Processes with Dependent Innovations. *Stochastic Processes and their Applications*, 115(6), 939–958.
- Yao, Y.-C., & Davis, R. A. (1986). The Asymptotic Behaviour of the Likelihood Ratio Statistic for Testing a Shift in Mean in a Sequence of Independent Normal Variables. *Sankhyā: The Indian Journal of Statistics Series A*, 48(3), 339–353.