Change detection problems in branching processes

Ph. D. Thesis

by

Tamás T. Szabó

Thesis advisor:

Professor Gyula Pap

Doctoral School of Mathematics and Computer Science, Bolyai Institute
University of Szeged, Faculty of Science and Informatics
Szeged, 2017

Acknowledgments

This thesis could never have been written without the help of a great many people.

First and foremost, I am eternally grateful to my advisor, Professor Gyula Pap, who has been an extraordinary mentor to me throughout the years, and even dealt with my occasional displays of stereotypical graduate-student behavior gracefully. He has constantly opened up new avenues of research (many of which are still unexplored), and accompanied me every step of the way. A glance at the references at the end of this thesis will convince the reader about the results of the school he has built over the years, which I am proud to be a student of.

Mátyás Barczy has been a very supportive co-author, and the incredible keenness of his eye has uncovered numerous mistakes, both large and small, in my manuscripts. The Bolyai Institute has been an outstanding place for study, and for starting my scientific career; I have received nothing but kind assistance from every teacher and colleague I had there. In particular, Gábor Szűcs has taught me most of what I know about the foundations of probability theory, and through his example in the teaching and practice of mathematics he has influenced me more in these matters than almost anyone else. I also owe thanks to my high school teachers of mathematics at the Radnóti Miklós high school in Szeged, who have helped me cement my inclination towards this beautiful profession.

I have never regarded the pursuit of the mathematical sciences as a lonely occupation. Throughout the years I spent on studying mathematics, I have been surrounded by a wonderful set of fellow students, who have constantly motivated me to become better at what I did – Kristóf Körmendi deserves a special mention here. The town of Szeged contained numerous good friends, with whom I have spent many a welcomely distracting (and improving) evening and weekend. Without them, to borrow a line from Gábor, this thesis would have been completed much earlier, but my life would be a dreadful bore.

I cannot imagine ever coming this far without my extended family. They have stood with me through the good times and the bad, and have always provided me with a stable foundation, which is essential in completing any enterprise. I am, and will always be, deeply indebted to my parents, who first kindled my intellectual curiosity, to my sister, and especially to my lovely wife, who has witnessed firsthand the emotional hardships that were unavoidable in the process of putting this thesis together, had to relieve me of my parental duties when deadlines were around the corner, and yet never wavered in her support. My dear sons have successfully prevented me from sleeping through late-night flashes of intuition, and cheered me up when the writing became too difficult.

I am blessed to have so many good people standing by me. One way or another, big or small, they all had a hand in this thesis, except for the errors – those are mine alone.

Contents

Chapter 1

Introduction

1.1 Historical background

Change detection is a naturally occurring question in statistics, and time series analysis in particular. One of the most widely used assumptions in time series analysis is that the dynamics of the process do not change over time, which allows us to collect a large enough samples for analysis. Obtaining a test for that assumption is therefore a natural desire.

The most widely cited early papers are Page (1954, 1955), and much of the early work was done in the field of control theory. The focus was at first (and to some extent, still is) on detecting a change in the mean of a series of independent variables. The distributions of the variables were often assumed to come from some parametric family, enabling the statistician to use likelihood methods to test for a change. Two main generalizations of that first model have suggested themselves from the outset: nonparametric cases, in which the null hypothesis only states that our sample consists of i.i.d. variables, and classical time series such as autoregressive moving average (ARMA) or generalized autoregressive conditional heteroskedasticity (GARCH) models, and more recently, even functional observations (Berkes et al., 2009).

As we have mentioned, likelihood-based methods have been successful for a wide range of processes. Most of these results are skilfully collected and presented in the canonical monograph of Csörgő and Horváth (1997). However, with our processes, the exact likelihood is usually unavailable because we do not make any distribution assumptions. That being said, our approach is motivated by the quasi-likelihood method of Gombay (2008).

The original cumulative sums (CUSUM) method of Page also continued to receive attention and was used in contexts where the likelihood function was either incalculable or impractical – see, e.g., Ploberger and Krämer (1992) and Lee et al. (2003). Also known as the Hinkley method, it was used for ARMA models by Baikovicius and Gerencsér (1992) and for hidden Markov models by Gerencsér and Prosdocimi (2010).

U-statistics have also been applied to the problem, especially if there were no distribution assumptions on the observations. For example, Gombay (2001) considered results under the alternative hypothesis as well.

Also, a significant portion of the literature is concerned with sequential or online methods, i.e., the observations are assumed to arrive one after another, and the objective is to detect a change as soon as possible. Berkes et al. (2004) used quasi-likelihood scores, closely related to our process, for this task. In contrast, our method will be offline, that is, we will receive all of our observations before starting the analysis.

The main objective of this thesis is to prove asymptotic properties of the testing procedure under the alternative hypothesis as well as the null hypothesis. We believe this to be important because, if investigated only under the null hypothesis, a change-detection procedure is essentially a model-fitting test, and results under the alternative are necessary to verify its use for the more special task of change detection.

The difficulty in applying the standard results to branching processes lies in the additional randomness introduced by the branching mechanism. This makes it impossible to follow likelihood-based methods except under very special circumstances. We are left, then, with few tools. Our method of choice is the martingale approach, since its foundations are well-explored and have several canonical monographs – Karatzas and Shreve (1991) and Jacod and Shiryaev (2003) will be the ones we use most. Also, martingale methods can be relatively easily extended to continuous time, as demonstrated below.

Nevertheless, we will not try to conceal the fact that our research was mainly motivated by Csörgő and Horváth (1997) and Gombay (2008), even if their methods of proof were not applicable to our processes. That motivation is largely explained by the fact that our first focus was the INAR(p) process, which resembles an AR(p) process in its covariance structure so it was hoped that the results for AR(p) processes can be extended to it. Indeed that hope has come true, and the following pages hopefully demonstrate how much further that idea can be taken if one is willing to perform more extensive calculations.

1.2 Notations

Let $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the sets of positive integers, non-negative integers, real numbers, non-negative real numbers and positive real numbers, respectively. For $x, y \in \mathbb{R}$, we will use $x \wedge y := \min(x, y)$. By $\|\boldsymbol{x}\|$, $\|\boldsymbol{A}\|$ and $\rho(\boldsymbol{A})$, we denote the Euclidean norm of a vector $\boldsymbol{x} \in \mathbb{R}^d$, the induced matrix norm of a matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$, and the spectral radius of $\boldsymbol{A}$, respectively. By $\boldsymbol{E}_d \in \mathbb{R}^{d\times d}$, we denote the $d \times d$ identity matrix (the more common notation $\boldsymbol{I}$ will be reserved for our information matrices), and by $\mathbf{1}_i$, the i-th unit vector. The Borel σ-algebra on $\mathbb{R}$ will be denoted by $\mathcal{B}(\mathbb{R})$. Continous martingales will make a frequent appearance; as usual, their quadratic variation will be denoted by $\langle\cdot\rangle$.

Unless otherwise noted, asymptotic statements are all to be understood as $T \to \infty$, or $n \to \infty$, as appropriate for continuous and discrete time, respectively. Modes of convergence will be stated as $\xrightarrow{\mathbb{P}}$, $\xrightarrow{\mathcal{D}}$ and $\xrightarrow{\text{a.s.}}$ for convergence in probability, in distribution and almost surely, respectively. For rates of convergence, we will use the Landau asymptotic notation: for a stochastic process X_t, the notation $X_t = \mathrm{O}_{\mathbb{P}}(g(t))$ means that the collection of measures $\big(\mathcal{L}\big(\frac{X_t}{g(t)}\big)\big)_{t>t_0}$ is tight for some $t_0 \geqslant 0$ ($\mathcal{L}$ stands for distribution, and note that t can be either discrete or continuous here). Also, $X_t = \mathrm{o}_{\mathbb{P}}(g(t))$ means simply that $\frac{X_t}{g(t)} \xrightarrow{\mathbb{P}} 0$.

As for the probabilistic setup, in continuous time, $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P}\big)$ will always be a filtered probability space satisfying the usual conditions, i.e., $(\Omega, \mathcal{F}, \mathbb{P})$ is complete, the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is right-continuous and $\mathcal{F}_0$ contains all the $\mathbb{P}$-null sets in $\mathcal{F}$. Also, in case of the Heston model, $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ will correspond to $(W_t, B_t)_{t\in\mathbb{R}_+}$, a standard Wiener process, and a given initial value (η_0, ζ_0), independent of $(W_t, B_t)_{t\in\mathbb{R}_+}$, such that $\mathbb{P}(\eta_0 \in \mathbb{R}_+) = 1$. Details of this construction can be found in Karatzas and Shreve (1991, Section 5.2).

1.3 Key results: change detection

Chapters 2 and 4 describe change detection tests for a discrete and a continuous time process, respectively. These tests have a lot of common steps, which are summarized below:

1. We will take a vector-valued process $\boldsymbol{X}_t$, indexed either by the natural numbers or the nonnegative real numbers and take a sample of it on the interval $0 \leqslant t \leqslant T$.

2. We will choose a parameter $\boldsymbol{\theta}_t$ governing the dynamics of the process. The main question will be whether this parameter is constant in t, or, formally, we would like to test

$$\mathrm{H}_0 : \ \exists \boldsymbol{\theta} : \ \boldsymbol{\theta}_t = \boldsymbol{\theta}, \quad t \in [0, T]$$

against the alternative hypothesis

$$\mathrm{H}_\mathrm{A} : \exists \rho \in (0,1) : \ \boldsymbol{\theta}_t = \boldsymbol{\theta}', \ t \in [0, \rho T) \ \text{ and } \ \boldsymbol{\theta}_t = \boldsymbol{\theta}'', \ t \in [\rho T, T]$$

for some $\boldsymbol{\theta}' \neq \boldsymbol{\theta}''$. An important additional condition will be for stability: $\boldsymbol{\theta}$, $\boldsymbol{\theta}'$, $\boldsymbol{\theta}''$ have to be such that $\boldsymbol{X}$ have a unique stationary distribution under H_0, and both parts of the process (before and after the change) have a unique stationary distribution under H_A.

3. We will find an appropriate vector-valued function f such that

$$\boldsymbol{M}_t := \boldsymbol{X}_t - \boldsymbol{X}_0 - \int_0^t f(\boldsymbol{\theta}_s; \boldsymbol{X}_{s-}) \, \mathrm{d}s$$

will be a martingale. Here $\boldsymbol{X}_{s-}$, a slightly informal notation, means $\boldsymbol{X}_s$ for continuous s and $\boldsymbol{X}_{s-1}$ for discrete s. Similarly, the integral is simply a sum for discrete s.

4. Assuming $\boldsymbol{\theta}_t = \boldsymbol{\theta}$ for all t, we will estimate $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}_T$ based on the conditional least squares (CLS) method of Klimko and Nelson (1978).

5. We will replace $\boldsymbol{\theta}_t$ with $\widehat{\boldsymbol{\theta}}_T$ in the definition of $\boldsymbol{M}_t$ to obtain $\widehat{\boldsymbol{M}}_t^{(T)}$.

6. We will prove that if $\boldsymbol{\theta}_t$ is constant in t, then

$$\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)} := \widehat{\boldsymbol{I}}_T^{-1/2} \widehat{\boldsymbol{M}}_{uT}^{(T)}, \quad u \in [0,1]$$

converges in distribution to a Brownian bridge on $[0,1]$, for some random normalizing matrix $\widehat{\boldsymbol{I}}_T$, which is calculable from the sample.

7. Consequently, we will construct tests for the change in $\boldsymbol{\theta}$, using the supremum or infimum of $\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)}$ as a test statistic (based on the direction of change).

8. We will prove that if there is a single change in $\boldsymbol{\theta}_t$ on $[0,T]$, then the test statistic will tend to infinity stochastically as $T \to \infty$.

9. We will prove that the *arg max*, or *arg min*, of $\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)}$ is a good estimator of the change point in $\boldsymbol{\theta}_t$.

Now we introduce the two special cases in our focus.

1.3.1 The INAR(p) process

The integer-valued autoregressive process of order p (denoted by INAR(p)) is defined by the following equation:

$$X_k = \alpha_1 \circ X_{k-1} + \cdots + \alpha_p \circ X_{k-p} + \varepsilon_k, \; k \in \mathbb{N}, \tag{1.3.1}$$

where the ε_k are i.i.d nonnegative integer-valued random variables with mean μ, and for a random nonnegative integer-valued random variable Y and $\alpha \in (0,1)$, $\alpha \circ Y$ denotes the sum of Y i.i.d Bernoulli random variables with mean α, also independent of Y. This process was first proposed by Alzaid and Al-Osh (1987) for $p = 1$ and Du and Li (1991) for higher p values. In this case the parameter vector will be

$$\boldsymbol{\theta} := \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ \mu \end{bmatrix}.$$

Change detection methods for INAR(p) processes in general (i.e., with no prespecified innovation distribution) have only been proposed in a few papers – we refer to Kang and Lee (2009) especially, where the authors give a test statistics similar to ours for a more general model. However, no result are available in these papers under the alternative hypothesis and the asymptotics of the change-point estimator are not given – we will give some answers to both of these questions which strengthen the theoretical foundations of the method considerably. Also, Hudecová et al. (2015) used the probability generating function to detect changes in the process. They prove the consistency of their test, but do not give the asymptotics of the change-point estimator.

We will define our martingale (or rather, its martingale differences) in (2.2.4), and the CLS estimates in (2.4.1). Our test process will be given in (2.5.2), and its asymptotic distribution under H_0 in Theorem 2.5.1. Under the alternative hypothesis the weak consistency of the test is proved in Theorem 2.7.1, and the asymptotic properties of the change-point estimator in Theorem 2.8.1. Finally, we will give a numerical illustration based on a widely analyzed polio dataset in 2.10. These results have been published in Pap and Szabó (2013). Appendix A contains some additional results, which our not necessary for our tests but are worthy of interest from a theoretical standpoint – in particular Theorem A.3.5 about the strong approximation of our test process under H_0.

1.3.2 The Cox–Ingersoll–Ross process

The Cox–Ingersoll–Ross (CIR) process:

$$\mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma\sqrt{Y_t}\,\mathrm{d}W_t, \quad t \in \mathbb{R}_+, \tag{1.3.2}$$

where $a \in \mathbb{R}_{++}$, $b \in \mathbb{R}_{++}$, $\sigma \in \mathbb{R}_{++}$ and $(W_t)_{t\in\mathbb{R}_+}$ is a standard Wiener process. The constraints on the parameter values ensure the ergodic behavior of our process – for details see Theorem 3.2.1 below. These constraints also ensure that any solution of (1.3.2) starting from a nonnegative value stays nonnegative almost surely – see Proposition 3.1.1.

This process was first investigated by Feller (1951), proposed as a short-term interest-rate model by Cox et al. (1985), and became one of the most widespread "short rate" models in financial mathematics.

Because of the central role that the process plays in financial mathematics, it has received considerable interest from statisticians, but mostly in the space of parameter estimation. Overbeck (1998) provided estimators based on continuous time observations, while the low-frequency discrete time CLS estimators were proposed by Overbeck and Rydén (1997). Li and Ma (2015) extended the investigation to so-called stable CIR processes driven by an α-stable process instead of a Brownian motion.

In change point detection, more regular processes (such as Ornstein–Uhlenbeck) and sequential analysis received most of the attention. The CIR process, which can be constructed as a limit of branching processes, presents a more challenging problem since many standard results are not directly applicable to it, chiefly because the diffusion coefficient is not Lipschitz continuous. Consequently, there are a handful of change detection tests for the CIR process in the literature: Schmid and Tzotchev (2004) used control charts and a sequential method i.e., an online procedure. In contrast, our approach is offline, where we assume the full sample to be known before starting investigations. They also supposed noisy observations, which will not be our interest. Guo and Härdle (2017) used the local parameter approach based on approximate maximum likelihood estimates. In essence, they wanted to find the largest interval for which the sample fits the model. Their method is based on a discrete sample, whereas we will use a continuous one.

It turns out that the approach outlined in Chapter 2 can be extended to the CIR process with relatively straightforward modifications, at least concerning the statement. Adapting the proofs, however, required some results that are noteworthy on their own. In particular, in Lemma 4.7.2, which will be essential in continuous time, we prove a Hájek-Rényi type result estimating the tail probabilities of the supremum of a continuous time stochastic process. This is an extension to continuous time of Lemma 2.9.1, and we believe it to be a new result that may find other applications beside this particular one.

Our parameter vector in this case will be

$$\boldsymbol{\theta} := \begin{bmatrix} a \\ b \end{bmatrix}$$

Change detection in σ is not necessary, since we can establish almost surely whether σ is constant across our sample. Indeed, the volatility parameter σ can be calculated almost surely from an arbitrarily small part of a continuous sample, see, e.g., Barczy and Pap (2016, Remark 2.6) or Overbeck and Rydén (1997, remark after Theorem 3.6).

We will give our estimators in (4.1.1), the martingale in (4.2.1), and the test process in (4.2.4). The asymptotic distribution of the test process under the null hypothesis will be given by Theorem 4.2.1. Weak consistence under the alternative hypothesis is proved in Theorem 4.5.1 and the asymptotic properties of the change-point estimator are given in Theorem 4.6.1. These results have been published in Pap and Szabó (2016).

1.4 Key results: parameter estimation

In the final chapter we will continue to extend our investigations by proposing conditional least squares estimators for the Heston model. The Heston model is a solution of a two-dimensional stochastic differential equation:

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma_1\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (\alpha - \beta Y_t)\,\mathrm{d}t + \sigma_2\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big), \end{cases} \qquad t \geqslant 0, \tag{1.4.1}$$

where $a \in \mathbb{R}_{++}$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1 \in \mathbb{R}_{++}$, $\sigma_2 \in \mathbb{R}_{++}$, $\varrho \in (-1,1)$, and $(W_t, B_t)_{t \geqslant 0}$ is a 2-dimensional standard Wiener process, see Heston (1993). It is immediately apparent that Y is just the Cox–Ingersoll–Ross process introduced in (1.3.2). Various interpretations of Y and X in financial mathematics are mentioned in, e.g., Hurn et al. (2013, Section 4).

Historically, most efforts have concentrated on parameter estimation for the CIR model only, and not the higher dimension Heston model. Specifically, Theorems 3.1 and 3.3 in Overbeck and Rydén (1997) correspond to our Theorem 5.4.2, but they estimate the volatility coefficient σ_1 as well, which we will assume to be known. For a more complete overview of parameter estimation for the Heston model see, e.g., the introduction in Barczy and Pap (2016).

We will focus on the subcritical case exclusively, i.e., when $b > 0$ (see Definition 3.1.3). In this case, just as in the previous section, the process Y has a unique stationary distribution. We would like to introduce conditional least squares estimators (CLSE's) for (a, b, α, β) based on discrete time observations. It will turn out, however, that, if not impossible, this is highly impractical, as the resulting partial derivatives depend on the parameters in a complicated manner. Therefore we transform the parameter space, and derive CLSE's for the transformed parameter vector, which will result in linear partial derivatives. Applying the inverse transformation to the CLSE's will lead to estimators for the original parameters, which could be considered CLSE's by a slight abuse of the term. However, we will refrain from referring to them as such.

We do not estimate the parameters σ_1, σ_2 and ϱ, since these parameters could – in principle, at least – be determined (rather than estimated) using an arbitrarily short continuous time observation $(X_t)_{t \in [0,T]}$ of X, where $T > 0$, see, e.g., Barczy and Pap (2016, Remark 2.6). In Overbeck and Rydén (1997, Theorems 3.2 and 3.3) one can find a strongly consistent and asymptotically normal estimator of σ_1 based on discrete time observations for the process Y. In any case, it will turn out that for the calculation of the estimator of (a, b, α, β), one does not need to know the values of the parameters σ_1, σ_2 and ϱ.

An alternative approach to using CLSE's would have been to calculate the discretized version of the maximum likelihood estimators derived in Barczy and Pap (2016) using the same procedure as Ben Alaya and Kebaier (2013, Section 4) apply for discrete time observations of high frequency. The relatively simple structure of the estimators we will arive at, however, makes them more appealing to us.

By this point it will hopefully become apparent how much easier the simple structure of CLSE's makes the task of change detection, an avenue which has been opened for the Heston model by these results, but not brought to its conclusion so far – this remains one of the most natural direction in which to extend our efforts. On the other hand, the lengthy calculations in the auxiliary lemmas in section 5.3 hint at the increasing complexity one has to face in advancing further towards constructing change detection procedures – even if that complexity promises to be merely computational and not conceptual.

Instead of the original parameters a, b, α, β, we will first estimate some transformed parameters, defined in (5.1.3). For these transformed parameters our estimates can be established by the standard CLS method in (5.1.4). The strong consistency and asymptotic normality of these estimators is proven in Theorem 5.2.1. After applying the inverse transformation, it is relatively easy to construct the estimates for the original values and obtain their asymptotic properties. The inverse transformation is given in (5.4.3), and the strong consistency and asymptotic normality of the resulting estimates are proven in Theorem 5.4.2. These results have been published in Barczy et al. (2016)

Chapter 2

The discrete case

2.1 Introductory definitions

A time-inhomogeneous INAR(p) process is a sequence $(X_k)_{k\geqslant -p+1}$ given by

$$X_k = \sum_{j=1}^{X_{k-1}} \xi_{1,k,j} + \cdots + \sum_{j=1}^{X_{k-p}} \xi_{p,k,j} + \varepsilon_k, \qquad k \in \mathbb{N}, \tag{2.1.1}$$

where $\{\varepsilon_k : k \in \mathbb{N}\}$ is a sequence of independent non-negative integer-valued random variables, for each $k \in \mathbb{N}$ and $i \in \{1, \ldots, p\}$ the sequence $\{\xi_{i,k,j} : j \in \mathbb{N}\}$ is a sequence of i.i.d. Bernoulli random variables with mean $\alpha_{i,k}$ such that these sequences are mutually independent and independent of the sequence $\{\varepsilon_k : k \in \mathbb{N}\}$, and $X_0, \ldots, X_{-p+1}$ are non-negative integer-valued random variables independent of the sequences $\{\xi_{i,k,j} : j \in \mathbb{N}\}$, $k \in \mathbb{N}$, $i \in \{1, \ldots, p\}$, and $\{\varepsilon_k : k \in \mathbb{N}\}$. The numbers $\alpha_{i,k}$ are called coefficients, and we will refer to $\varepsilon_1, \varepsilon_2, \ldots$ as the innovations. Time-homogeneous INAR(p) processes have a number of applications, which are summarized, e.g., in Barczy et al. (2011).

The reason that we initially define our process as time-inhomogeneous is that we would like to test for a change in the parameters, therefore we have to allow them to vary over time. In the proofs, however, a majority of the results will be based upon the properties of time-homogeneous INAR(p) processes.

Now we proceed with the formulation of the statistical problem. We assume $\mu_k := \mathbb{E}(\varepsilon_k) < \infty$ and $0 < \sigma_k^2 := \operatorname{Var}(\varepsilon_k) < \infty$. Write the parameter vectors as

$$\begin{bmatrix} \alpha_{1,k} \\ \vdots \\ \alpha_{p,k} \\ \mu_k \end{bmatrix} =: \boldsymbol{\theta}_k,$$

and let us choose a subset PV of $\{1, 2, \ldots, p+1\}$ such that

(2.1.2) $$PV = \{i_1, i_2, \ldots, i_\ell\} \text{ for some } \ell > 0, \quad i_1 < i_2 < \ldots < i_\ell.$$

Also, we can write

$$NV := PV^{\complement} = \{j_1, j_2, \ldots, j_{p+1-\ell}\}, \quad j_1 < j_2 < \ldots < j_{p+1-\ell},$$

where $H^{\complement}$ denotes the complement of a set. Let us now define

$$\boldsymbol{\varphi}_k := (\theta_k^{(i_1)}, \theta_k^{(i_2)}, \ldots, \theta_k^{(i_\ell)})^\top, \quad \boldsymbol{\eta}_k := (\theta_k^{(j_1)}, \theta_k^{(j_2)}, \ldots, \theta_k^{(j_{p+1-\ell})})^\top.$$

The vector $\boldsymbol{\varphi}_k$ is the parameter vector of interest and $\boldsymbol{\eta}_k$ is the 'nuisance' parameter vector. For a fixed number of observations n we want to test the null hypothesis

$$\mathrm{H}_0: \qquad \varepsilon_1, \ldots, \varepsilon_n \text{ are identically distributed and } \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \cdots = \boldsymbol{\theta}_n$$

against the alternative

$$\begin{aligned} \mathrm{H}_\mathrm{A}: \qquad & \text{there is an integer } \tau \in \{1, \ldots, n-1\} \text{ such that} \\ & \boldsymbol{\varphi}_1 = \cdots = \boldsymbol{\varphi}_\tau \neq \boldsymbol{\varphi}_{\tau+1} = \cdots = \boldsymbol{\varphi}_n \text{ but } \boldsymbol{\eta}_1 = \cdots = \boldsymbol{\eta}_n, \\ & \varepsilon_1, \ldots, \varepsilon_\tau \text{ are identically distributed,} \\ & \text{and } \varepsilon_{\tau+1}, \ldots, \varepsilon_n \text{ are identically distributed.} \end{aligned}$$

Under the null hypothesis H_0, then, we have

$$\begin{bmatrix} \alpha_{1,1} \\ \vdots \\ \alpha_{p,1} \\ \mu_1 \end{bmatrix} = \cdots = \begin{bmatrix} \alpha_{1,n} \\ \vdots \\ \alpha_{p,n} \\ \mu_n \end{bmatrix} =: \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ \mu \end{bmatrix} =: \begin{bmatrix} \boldsymbol{\alpha} \\ \mu \end{bmatrix} =: \boldsymbol{\theta}, \qquad \sigma_1^2 = \ldots = \sigma_n^2 =: \sigma^2.$$

2.2 Regression equations

The INAR(p) process is formally analogous to the AR(p) process. To exploit this analogy we need to state several regression equations for the process.

First we create a Markov chain from our process the usual way, by extending the state space:

$$\boldsymbol{X}_k := \begin{bmatrix} X_k \\ X_{k-1} \\ \vdots \\ X_{k-p} \end{bmatrix}. \tag{2.2.1}$$

The equivalent of (2.1.1) for the vector-valued process $(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ defined in (2.2.1) is

$$\boldsymbol{X}_k = \sum_{i=1}^{p} \sum_{j=1}^{X_{k-i}} \boldsymbol{\xi}_{i,k,j} + \boldsymbol{\varepsilon}_k, \tag{2.2.2}$$

where

$$\boldsymbol{\xi}_{1,k,j} = \begin{bmatrix} \xi_{1,k,j} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \boldsymbol{\xi}_{2,k,j} = \begin{bmatrix} \xi_{2,k,j} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ \dots, \ \boldsymbol{\xi}_{p-1,k,j} = \begin{bmatrix} \xi_{p-1,k,j} \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \ \boldsymbol{\xi}_{p,k,j} = \begin{bmatrix} \xi_{p,k,j} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \boldsymbol{\varepsilon}_k = \begin{bmatrix} \varepsilon_k \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This form makes it even more apparent that the INAR(p) process is a special multi-type branching process with immigration. According to standard literature (see, e.g., Quine, 1970), if the matrix

$$\boldsymbol{A} := \begin{bmatrix} \mathbb{E}(\boldsymbol{\xi}_{1,1,1}) & \cdots & \mathbb{E}(\boldsymbol{\xi}_{p,1,1}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{2.2.3}$$

is primitive (i.e., some power of it is elementwise positive), the ergodicity of the process depends only on the spectral radius $\rho(A)$, and the process is ergodic if $\rho(A) < 1$. In Barczy et al. (2011, (2.7)), it is shown that this is equivalent to the condition that $\alpha_1 + \ldots + \alpha_p < 1$. For the primitivity, it is sufficient to have $\alpha_p > 0$ and that the greatest common denominator of the numbers i such that $\alpha_i > 0$ is 1. Equivalently, the process (2.1.1) is referred to as stable, unstable or explosive whenever $\alpha_1 + \cdots + \alpha_p < 1$, $\alpha_1 + \cdots + \alpha_p = 1$ or $\alpha_1 + \cdots + \alpha_p > 1$, respectively. Basic differences between the three types are summarized in Barczy et al. (2011). In this terminology, we will study only the stable case $\alpha_1 + \cdots + \alpha_p < 1$ with $\alpha_p > 0$. To condense these conditions into one, we introduce the following definition.

2.2.1 Definition. *A time-homogeneous INAR(p) process* $(X_k)_{k\geqslant -p+1}$ *is said to satisfy condition* $\mathbf{C}_0$*, if* $\mathbb{E}(X_0^6) < \infty,\ \ldots,\ \mathbb{E}(X_{-p+1}^6) < \infty,\ \ \mathbb{E}(\varepsilon_1^6) < \infty,\ \ \alpha_1 + \cdots + \alpha_p < 1,\ \ \mu > 0$ *all hold for it, and if, furthermore,* $\alpha_p > 0$ *and the greatest common denominator of the numbers* i *such that* $\alpha_i > 0$ *is 1.*

To further emphasize the similarities with AR(p) processes, we can define

$$M_k = X_k - \mathbb{E}(X_k|\mathcal{F}_{k-1}) = X_k - \boldsymbol{\alpha}^\top \boldsymbol{X}_{k-1} - \mu, \qquad k \in \mathbb{N}, \tag{2.2.4}$$

where $(\mathcal{F}_n)_{n\in\mathbb{N}}$ is the natural filtration, and write

$$\boldsymbol{X}_k = \boldsymbol{A}\boldsymbol{X}_{k-1} + (\mu + M_k)\mathbf{1}_1, \tag{2.2.5}$$

where $\mathbf{1}_1$ is the first unit vector. It is clear from the definition that $M_k,\ k \in \mathbb{N}$ is a series of martingale differences. Based on (2.2.5) we obtain

$$\begin{aligned}
\boldsymbol{X}_k^{\otimes 2} &= (\boldsymbol{A}\boldsymbol{X}_{k-1})^{\otimes 2} + ((\mu + M_k)\mathbf{1}_1)^{\otimes 2} + (\boldsymbol{A}\boldsymbol{X}_{k-1}) \otimes ((\mu + M_k)\mathbf{1}_1) \\
&\quad + ((\mu + M_k)\mathbf{1}_1) \otimes (\boldsymbol{A}\boldsymbol{X}_{k-1}) \\
&= \boldsymbol{A}^{\otimes 2}\boldsymbol{X}_{k-1}^{\otimes 2} + (\mu + M_k)^2 \mathbf{1}_1^{\otimes 2} + (\mu + M_k)(\boldsymbol{A}\boldsymbol{X}_{k-1}) \otimes \mathbf{1}_1 \\
&\quad + (\mu + M_k)\mathbf{1}_1 \otimes (\boldsymbol{A}\boldsymbol{X}_{k-1}),
\end{aligned} \tag{2.2.6}$$

where $\otimes$ denotes Kronecker product of matrices.

2.2.2 Remark. We have chosen the INAR(p) process on account of its ubiquity in modelling integer valued time series. However, in the following considerations, nothing about the matrix $\boldsymbol{A}$ will be exploited other than its spectral radius. Therefore, our results should be equally applicable to general p-type branching processes as well.

2.3 Asymptotic properties of the process under $\mathbf{C}_0$

Under $\mathbf{C}_0$ let us denote by $\widetilde{\boldsymbol{X}}$ a random vector with the unique stationary distribution of $(\boldsymbol{X}_k)_{k\geqslant -p+1}$. Because our process is ergodic, we can apply the ergodic theorem. In its well-known form it states that if $\mathbb{E}(|g(\widetilde{\boldsymbol{X}})|) < \infty$ for some function g, then

$$\frac{1}{n}\sum_{k=1}^{n} g(\boldsymbol{X}_k) \xrightarrow{\text{a.s.}} \mathbb{E}(g(\widetilde{\boldsymbol{X}})). \tag{2.3.1}$$

This is, for example, Theorem 2 in I.15. in Chung (1960). However, instead of the convergence of averages, we will frequently require the convergence of expectations, i.e.,

$$\mathbb{E}(\boldsymbol{X}_k^{\otimes\beta}) \to \mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes\beta}), \qquad \beta \in \mathbb{N}, \tag{2.3.2}$$

whenever the right hand side is finite. This is Theorem 14.0.1 in Meyn and Tweedie (2009). The result in (2.3.2) also implies convergence of any component of the matrices. Under the alternative hypothesis we will additionally apply

$$\sum_{x \in \mathbb{Z}_+^p} |\,\mathbb{P}(\boldsymbol{X}_n = x) - \mathbb{P}(\widetilde{\boldsymbol{X}} = x)| \to 0. \tag{2.3.3}$$

This result can be found in Meyn and Tweedie (2009, Theorem 13.1.2). We conclude our introductory remarks with a definition. The following vector contains the variances of the Bernoulli variables used in the evolution of the process:

$$\boldsymbol{\alpha}_* := [\alpha_1(1-\alpha_1), \ldots, \alpha_p(1-\alpha_p)]^\top. \tag{2.3.4}$$

The convergence rate in (2.3.2) can be estimated by the following lemma.

2.3.1 Lemma. *Under* $\mathbf{C}_0$ *there is a constant* $\pi \in (0,1)$ *such that*

$$\|\,\mathbb{E}(\boldsymbol{X}_k) - \mathbb{E}(\widetilde{\boldsymbol{X}})\| = \mathrm{O}(\pi^k), \qquad \|\,\mathbb{E}(\boldsymbol{X}_k^{\otimes 2}) - \mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes 2})\| = \mathrm{O}(\pi^k).$$

Proof.

We use (2.2.5) to conclude that

$$\mathbb{E}(\boldsymbol{X}_k) = \boldsymbol{A}\,\mathbb{E}(\boldsymbol{X}_{k-1}) + \mu\mathbf{1}_1.$$

Taking the limits as $k \to \infty$ we have

$$\mathbb{E}(\widetilde{\boldsymbol{X}}) = \boldsymbol{A}\,\mathbb{E}(\widetilde{\boldsymbol{X}}) + \mu\mathbf{1}_1,$$

hence

$$\mathbb{E}(\boldsymbol{X}_k) - \mathbb{E}(\widetilde{\boldsymbol{X}}) = \boldsymbol{A}(\mathbb{E}(\boldsymbol{X}_{k-1}) - \mathbb{E}(\widetilde{\boldsymbol{X}})). \tag{2.3.5}$$

Similarly, from (2.2.6) and $\mathbb{E}(M_k^2|\mathcal{F}_{k-1}) = \boldsymbol{\alpha}_*^\top \boldsymbol{X}_{k-1} + \sigma^2$ we have

(2.3.6)
$$\begin{aligned}
\mathbb{E}(\boldsymbol{X}_k^{\otimes 2}) - \mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes 2}) &= \boldsymbol{A}^{\otimes 2}(\mathbb{E}(\boldsymbol{X}_{k-1}^{\otimes 2}) - \mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes 2})) + \big(\boldsymbol{\alpha}_*^\top(\mathbb{E}(\boldsymbol{X}_{k-1}) - \mathbb{E}(\widetilde{\boldsymbol{X}}))\big)\mathbf{1}_1^{\otimes 2} \\
&\quad + \mu(\boldsymbol{A}(\mathbb{E}(\boldsymbol{X}_{k-1}) - \mathbb{E}(\widetilde{\boldsymbol{X}}))) \otimes \mathbf{1}_1 + \mu\mathbf{1}_1 \otimes (\boldsymbol{A}(\mathbb{E}(\boldsymbol{X}_{k-1}) - \mathbb{E}(\widetilde{\boldsymbol{X}}))) \\
&= \boldsymbol{A}^{\otimes 2}(\mathbb{E}(\boldsymbol{X}_{k-1}^{\otimes 2}) - \mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes 2})) + \mathbf{1}_1^{\otimes 2}\boldsymbol{\alpha}_*^\top(\mathbb{E}(\boldsymbol{X}_{k-1}) - \mathbb{E}(\widetilde{\boldsymbol{X}})) \\
&\quad + \mu(\boldsymbol{A} \otimes \mathbf{1}_1)(\mathbb{E}(\boldsymbol{X}_{k-1}) - \mathbb{E}(\widetilde{\boldsymbol{X}})) + \mu(\mathbf{1}_1 \otimes \boldsymbol{A})(\mathbb{E}(\boldsymbol{X}_{k-1}) - \mathbb{E}(\widetilde{\boldsymbol{X}})).
\end{aligned}$$

Here we used the fact that for any $c \in \mathbb{R}$ and real vector $\boldsymbol{v}$ we have $c\boldsymbol{v} = \boldsymbol{v}c$, where the second multiplication is a proper matrix product. Furthermore, we used the following property of the Kronecker product: for any matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$, if the operations on both sides are permitted, we have $(\boldsymbol{AB}) \otimes (\boldsymbol{CD}) = (\boldsymbol{A} \otimes \boldsymbol{C})(\boldsymbol{B} \otimes \boldsymbol{D})$; specifically, if $\boldsymbol{C}$ is a column vector, $(\boldsymbol{AB}) \otimes \boldsymbol{C} = (\boldsymbol{AB}) \otimes (\boldsymbol{C} \cdot 1) = (\boldsymbol{A} \otimes \boldsymbol{C})(\boldsymbol{B} \otimes 1) = (\boldsymbol{A} \otimes \boldsymbol{C})\boldsymbol{B}$ (this identity can also be used when the first factor consists of a single factor instead of the second). Hence,

$$\begin{bmatrix} \mathbb{E}(\boldsymbol{X}_k) - \mathbb{E}(\widetilde{\boldsymbol{X}}) \\ \mathbb{E}(\boldsymbol{X}_k^{\otimes 2}) - \mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes 2}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{1}_1^{\otimes 2}\boldsymbol{\alpha}_*^\top + \mu(\boldsymbol{A} \otimes \boldsymbol{1}_1) + \mu(\boldsymbol{1}_1 \otimes \boldsymbol{A}) & \boldsymbol{A}^{\otimes 2} \end{bmatrix} \begin{bmatrix} \mathbb{E}(\boldsymbol{X}_{k-1}) - \mathbb{E}(\widetilde{\boldsymbol{X}}) \\ \mathbb{E}(\boldsymbol{X}_{k-1}^{\otimes 2}) - \mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes 2}) \end{bmatrix}.$$

Let us denote the multiplicating matrix on the right hand side by $\boldsymbol{D}$. We note that $\boldsymbol{D}$ is block lower triangular and that due to the properties of the Kronecker product, $\rho(\boldsymbol{A}^{\otimes 2}) = (\rho(\boldsymbol{A}))^2 < \rho(\boldsymbol{A})$. From these it is clear that $\rho(\boldsymbol{D}) = \rho(\boldsymbol{A}) < 1$. It is well-known that then there exists an induced matrix norm $\|\cdot\|_*$ for which $\rho(\boldsymbol{A}) < \|\boldsymbol{A}\|_* < 1$. This, and the equivalence of vector norms suffice for the proof. □

2.3.2 Remark. The finiteness of the respective moments of the stationary distribution can be derived using the same approach as in the proof of formulas (2.2.3), (2.2.4) and (2.2.10) in Barczy et al. (2011). We note that the stationary distribution has exactly as many finite moments as the innovation distribution and the initial distributions have in common, because the Bernoulli distribution is bounded and therefore all of its moments are finite.

The following lemma will be used for our results repeatedly. It shows that while the values of the process are not independent, their dependence is weak and the autocovariance of the process decays rapidly, so that the sum of all autocovariances up to n is linear in n.

2.3.3 Lemma. *Under* $\mathbf{C}_0$ *we have*

(i) $\mathrm{Var}(X_1 + X_2 + \ldots + X_n) = \sum_{i,j=1}^n \mathrm{Cov}(X_i, X_j) = \mathrm{O}(n)$,

(ii) $\mathrm{Var}(X_1X_{1-q} + X_2X_{2-q} + \ldots + X_nX_{n-q}) = \sum_{i,j=1}^n \mathrm{Cov}(X_iX_{i-q}, X_jX_{j-q}) = \mathrm{O}(n)$ *for all* $0 \leqslant q \leqslant p-1$.

Proof. Although the lemma is stated for the process $(X_n)_{n\in\mathbb{N}}$, calculations will require that we investigate the process $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$. Therefore, we will prove the following statements:

$$\|\mathrm{Var}(\boldsymbol{X}_1 + \boldsymbol{X}_2 + \ldots + \boldsymbol{X}_n)\| = \mathrm{O}(n) \tag{2.3.7}$$

in the place of (i) and

$$\left\|\mathrm{Var}(\boldsymbol{X}_1^{\otimes 2} + \boldsymbol{X}_2^{\otimes 2} + \ldots + \boldsymbol{X}_n^{\otimes 2})\right\| = \mathrm{O}(n) \tag{2.3.8}$$

in the place of (ii).

First we will prove (2.3.7). Since (2.3.2) implies that $(\|\mathrm{Var}(\boldsymbol{X}_i)\|)_{i\in\mathbb{N}}$ is a (convergent and hence) bounded series, we will only need to deal with

$$\sum_{i,j=1}^{n} \mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) - \sum_{i=1}^{n} \mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_i),$$

since the latter sum is clearly $\mathrm{O}(n)$. It is immediate from (2.2.5) that

$$\boldsymbol{X}_k - \mathbb{E}(\boldsymbol{X}_k) = \boldsymbol{A}(\boldsymbol{X}_{k-1} - \mathbb{E}(\boldsymbol{X}_{k-1})) + M_k \mathbf{1}_1.$$

Let us now fix $1 \leqslant i < j$ and write

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) &= \mathbb{E}\left[(\boldsymbol{X}_i - \mathbb{E}(\boldsymbol{X}_i))(\boldsymbol{X}_j - \mathbb{E}(\boldsymbol{X}_j))^\top\right] \\ &= \mathbb{E}\left[(\boldsymbol{X}_i - \mathbb{E}(\boldsymbol{X}_i))\,\mathbb{E}(\boldsymbol{X}_j - \mathbb{E}(\boldsymbol{X}_j)|\mathcal{F}_{j-1})^\top\right] \\ &= \mathbb{E}\left[(\boldsymbol{X}_i - \mathbb{E}(\boldsymbol{X}_i))(\boldsymbol{A}(\boldsymbol{X}_{j-1} - \mathbb{E}(\boldsymbol{X}_{j-1})))^\top\right] = \mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_{j-1})\boldsymbol{A}^\top. \end{aligned} \tag{2.3.9}$$

If we perform the calculations for the case $1 \leqslant j < i$ as well, we can see that, after $|j-i|$ iterations,

$$\mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \boldsymbol{A}^{(i-j)_+}\,\mathrm{Var}(\boldsymbol{X}_{\min(i,j)})(\boldsymbol{A}^\top)^{(j-i)_+}. \tag{2.3.10}$$

Because $\rho(\boldsymbol{A}) < 1$, there exists a matrix norm $\|\cdot\|_*$ for which $\rho(\boldsymbol{A}) < \|\boldsymbol{A}\|_* =: \pi < 1$. With this norm, (2.3.10), and the boundedness of $(\mathrm{Var}(\boldsymbol{X}_i))_{i\in\mathbb{N}}$ we can establish

$$\|\mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_j)\|_* = \mathrm{O}(\pi^{|i-j|}),$$

which yields (2.3.7) immediately.

For (2.3.8) our reasoning will be very similar, although with more tedious calculations. First we note that (2.3.2) implies boundedness for $\left\|\mathrm{Var}(\boldsymbol{X}_i^{\otimes 2})\right\|$ also. From (2.2.6) we have

$$\begin{aligned} \mathbb{E}(\boldsymbol{X}_k^{\otimes 2} - \mathbb{E}(\boldsymbol{X}_k^{\otimes 2})|\mathcal{F}_{k-1}) &= \boldsymbol{A}^{\otimes 2}(\boldsymbol{X}_{k-1}^{\otimes 2} - \mathbb{E}(\boldsymbol{X}_{k-1}^{\otimes 2})) + \boldsymbol{\alpha}_*^\top(\boldsymbol{X}_{k-1} - \mathbb{E}(\boldsymbol{X}_{k-1}))\mathbf{1}_1^{\otimes 2} \\ &\quad + \mu[\boldsymbol{A}(\boldsymbol{X}_{k-1} - \mathbb{E}(\boldsymbol{X}_{k-1}))] \otimes \mathbf{1}_1 + \mu \mathbf{1}_1 \otimes [\boldsymbol{A}(\boldsymbol{X}_{k-1} - \mathbb{E}(\boldsymbol{X}_{k-1}))] \\ &= \boldsymbol{A}^{\otimes 2}(\boldsymbol{X}_{k-1}^{\otimes 2} - \mathbb{E}(\boldsymbol{X}_{k-1}^{\otimes 2})) + \mathbf{1}_1^{\otimes 2}\boldsymbol{\alpha}_*^\top(\boldsymbol{X}_{k-1} - \mathbb{E}(\boldsymbol{X}_{k-1})) \\ &\quad + \mu[\boldsymbol{A} \otimes \mathbf{1}_1](\boldsymbol{X}_{k-1} - \mathbb{E}(\boldsymbol{X}_{k-1})) + \mu(\mathbf{1}_1 \otimes \boldsymbol{A})(\boldsymbol{X}_{k-1} - \mathbb{E}(\boldsymbol{X}_{k-1})), \end{aligned}$$

analogously to (2.3.6). Now, similarly to (2.3.9) we get, for $1 \leqslant i < j$,

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_j^{\otimes 2}) &= \mathrm{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_{j-1}^{\otimes 2})(\boldsymbol{A}^{\otimes 2})^\top + \mathrm{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_{j-1})\boldsymbol{\alpha}_*(\mathbf{1}_1^{\otimes 2})^\top \\ &\quad + \mu\,\mathrm{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_{j-1})\,(\boldsymbol{A} \otimes \mathbf{1}_1)^\top + \mu\,\mathrm{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_{j-1})\,(\mathbf{1}_1 \otimes \boldsymbol{A})^\top. \end{aligned}$$

Here

$$\operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_{j-1}) := \mathbb{E}[(\boldsymbol{X}_i^{\otimes 2} - \mathbb{E}(\boldsymbol{X}_i^{\otimes 2}))(\boldsymbol{X}_{j-1} - \mathbb{E}(\boldsymbol{X}_{j-1}))^\top],$$

a $p^2 \times p$ matrix. Also similarly to (2.3.9) we have

$$\operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_j) = \operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_{j-1})\boldsymbol{A}^\top.$$

Summarizing, we get the following regression:
(2.3.11)

$$\begin{bmatrix} \operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_j)^\top \\ \operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_j^{\otimes 2})^\top \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{1}_1^{\otimes 2}\boldsymbol{\alpha}_*^\top + \mu(\boldsymbol{A} \otimes \boldsymbol{1}_1) + \mu(\boldsymbol{1}_1 \otimes \boldsymbol{A}) & \boldsymbol{A}^{\otimes 2} \end{bmatrix} \begin{bmatrix} \operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_{j-1})^\top \\ \operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_{j-1}^{\otimes 2})^\top \end{bmatrix}.$$

Note that the multiplicating matrix on the right hand side is just $\boldsymbol{D}$ from the proof of Lemma 2.3.1. Now, similarly to (2.3.10), we have

$$\begin{bmatrix} \operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_j)^\top \\ \operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_j^{\otimes 2})^\top \end{bmatrix} = \boldsymbol{D}^{(j-i)_+} \begin{bmatrix} \operatorname{Cov}(\boldsymbol{X}_{\min(i,j)}^{\otimes 2}, \boldsymbol{X}_{\min(i,j)})^\top \\ \operatorname{Cov}(\boldsymbol{X}_{\min(i,j)}^{\otimes 2}, \boldsymbol{X}_{\min(i,j)}^{\otimes 2})^\top \end{bmatrix} \left(\boldsymbol{D}^\top\right)^{(i-j)_+}. \tag{2.3.12}$$

Now we only need to note that $(\operatorname{Cov}(\boldsymbol{X}_i^{\otimes 2}, \boldsymbol{X}_i))_{i \in \mathbb{N}}$ is a bounded sequence due to (2.3.2), and we can finish the proof of (2.3.8) in the same way as (2.3.7). □

2.4 Conditional least squares estimates

Recalling (2.2.4), the conditional least squares estimators of the parameters, first introduced by Klimko and Nelson (1978), can be calculated by minimizing the sum of squares

$$R_n(\alpha_1, \ldots, \alpha_p, \mu) := \frac{1}{2}\sum_{k=1}^n M_k^2 = \frac{1}{2}\sum_{k=1}^n (X_k - \boldsymbol{\alpha}^\top \boldsymbol{X}_{k-1} - \mu)^2$$

with respect to $\alpha_1, \ldots, \alpha_p, \mu$. With a reasoning completely analogous to that of Lemma 3.1 and Proposition 3.1 in Barczy et al. (2014b) we can show that whenever $\boldsymbol{Q}_n$ is invertible, R_n has a unique minimum given by

$$\widehat{\boldsymbol{\theta}}_n := \begin{bmatrix} \widehat{\boldsymbol{\alpha}}_n \\ \widehat{\mu}_n \end{bmatrix} := \boldsymbol{Q}_n^{-1} \sum_{k=1}^n X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}, \qquad \boldsymbol{Q}_n := \sum_{k=1}^n \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top. \tag{2.4.1}$$

Also, with the help of (2.3.1),

$$\frac{\boldsymbol{Q}_n}{n} \xrightarrow{\text{a.s.}} \boldsymbol{Q} := \mathbb{E}\left(\begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix}^\top \right). \tag{2.4.2}$$

Now, we will show in A.1 that $\boldsymbol{Q}_n$ is, in fact, invertible with an asymptotic probability of 1, therefore the parameter estimates exist and are unique with an asymptotic probability of 1. As all our results our asymptotic, this will be sufficient for the purposes of the paper. Replacing the parameters by their estimates in M_k we obtain $\widehat{M}_k^{(n)}$, i.e.,

$$\widehat{M}_k^{(n)} := X_k - \widehat{\boldsymbol{\theta}}_n^\top \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}. \tag{2.4.3}$$

Although not a parameter in which we are looking for change, the estimate of the variance of the innovation σ^2 will also appear in our test process, therefore we have to provide an estimator for it. To do this, we introduce

$$N_k = M_k^2 - E(M_k^2|\mathcal{F}_{k-1}) = M_k^2 - \alpha_1(1-\alpha_1)X_{k-1} - \cdots - \alpha_p(1-\alpha_p)X_{k-p} - \sigma^2, \qquad k \geqslant 0.$$

Minimizing $\sum_{k=1}^n N_k^2$ with respect to σ^2 we obtain the conditional least squares estimate

$$\overline{\sigma}_n^2 = -\frac{1}{n}\sum_{k=1}^n (M_k^2 - \alpha_1(1-\alpha_1)X_{k-1} - \cdots - \alpha_p(1-\alpha_p)X_{k-p}). \tag{2.4.4}$$

However, in this estimate the true parameters are still present. The estimate that we will use is given by replacing the α coefficients and μ both in the formula and in M_k^2 by their estimates:

$$\widehat{\sigma}_n^2 = -\frac{1}{n}\sum_{k=1}^n \left(\left(\widehat{M}_k^{(n)}\right)^2 - \widehat{\alpha}_1^{(n)}\left(1-\widehat{\alpha}_1^{(n)}\right)X_{k-1} - \cdots - \widehat{\alpha}_p^{(n)}\left(1-\widehat{\alpha}_p^{(n)}\right)X_{k-p} \right). \tag{2.4.5}$$

The strong consistency of the estimates is established below. This is a well known result by Du and Li (1991), we only include it to demonstrate a relatively simple proof with our notations.

2.4.1 Theorem. (Du and Li, 1991) *If the process satisfies* $\mathbf{C}_0$ *then*

$$\widehat{\boldsymbol{\theta}}^{(n)} \xrightarrow{\text{a.s.}} \boldsymbol{\theta} \quad \textit{and} \quad \widehat{\sigma}_n^2 \xrightarrow{\text{a.s.}} \sigma^2. \tag{2.4.6}$$

Proof. First of all, it is a matter of simple calculations (see A.2) and a straightforward application of (2.3.1) that, provided that the second moment of ε_1 exists,

$$\frac{1}{n}\sum_{k=1}^n M_k^2 \to \boldsymbol{\alpha}_*^\top\, \mathbb{E}(\widetilde{\boldsymbol{X}}) + \sigma^2, \tag{2.4.7}$$

with $\boldsymbol{\alpha}_*$ given in (2.3.4). Hence, taking the limits of the expectations in (2.2.5) and (2.2.6) we have

$$\begin{aligned}\mathbb{E}(\widetilde{\boldsymbol{X}}) &= \boldsymbol{A}\,\mathbb{E}(\widetilde{\boldsymbol{X}}) + \mu\\ \mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes 2}) &= \boldsymbol{A}^{\otimes 2}\,\mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes 2}) + (\mu^2 + \boldsymbol{\alpha}_*^\top\,\mathbb{E}(\widetilde{\boldsymbol{X}}) + \sigma^2) + \mu(\boldsymbol{e}_1 \otimes \boldsymbol{A}\,\mathbb{E}(\widetilde{\boldsymbol{X}}) + \boldsymbol{A}\,\mathbb{E}(\widetilde{\boldsymbol{X}}) \otimes \boldsymbol{e}_1).\end{aligned}$$

Now we note that

$$\boldsymbol{U}_k := \begin{bmatrix} X_k \\ \boldsymbol{X}_{k-1} \end{bmatrix}, \quad k \in \mathbb{N}$$

satisfies a similar recursion to (2.2.2). The equivalent of the matrix $\boldsymbol{A}$ can then be shown to have a spectral radius smaller than 1, hence $(\boldsymbol{U}_k)_{k\in\mathbb{N}}$ is ergodic, and we can apply (2.3.1). Moreover, it is clear that if $\widetilde{\boldsymbol{U}}$ denotes a vector with the unique stationary distribution of $(\boldsymbol{U}_k)_{k\in\mathbb{N}}$ then

$$\begin{bmatrix} \widetilde{U}^{(2)} \\ \widetilde{U}^{(3)} \\ \vdots \\ \widetilde{U}^{(p+1)} \end{bmatrix} \stackrel{\mathcal{L}}{=} \widetilde{\boldsymbol{X}}$$

and for the components of $\widetilde{\boldsymbol{U}}$ we also have

$$\widetilde{U}^{(1)} \stackrel{\mathcal{L}}{=} \sum_{j=1}^{\widetilde{U}^{(2)}} \xi_{1,j} + \cdots + \sum_{j=1}^{\widetilde{U}^{(p+1)}} \xi_{p,j} + \varepsilon,$$

where $\xi_{i,j} \stackrel{\mathcal{L}}{=} \xi_{i,1,1}, i = 1, \ldots, p,\ j \in \mathbb{N}$ and $\varepsilon \stackrel{\mathcal{L}}{=} \varepsilon_1$ such that all these variables are totally independent and also independent of $(\widetilde{U}^{(2)}, \ldots, \widetilde{U}^{(p+1)})^\top$. Hence (2.3.1) implies.

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} &\xrightarrow{\mathbb{P}} \mathbb{E}\left(\left(\sum_{j=1}^{\widetilde{X}_0} \xi_{1,j} + \cdots + \sum_{j=1}^{\widetilde{X}_{-p+1}} \xi_{p,j} + \varepsilon\right)\begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix}\right)\\ &= \mathbb{E}\left(\left(\alpha_1 \widetilde{X}_0 + \cdots + \alpha_p \widetilde{X}_{-p+1} + \mu\right)\begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix}\right) = \boldsymbol{Q}\boldsymbol{\theta}\end{aligned} \tag{2.4.8}$$

as $n \to \infty$ (here $\varepsilon \stackrel{\mathcal{L}}{=} \varepsilon_1$).

From this we immediately conclude

$$\widehat{\boldsymbol{\theta}}^{(n)} = \left(\frac{\boldsymbol{Q}_n}{n}\right)^{-1} \frac{1}{n}\sum_{k=1}^{n} X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \xrightarrow{\text{a.s.}} \boldsymbol{Q}^{-1}\boldsymbol{Q}\boldsymbol{\theta} = \boldsymbol{\theta}.$$

A similar result can be derived for the estimate of σ^2. By recalling (2.4.7) and computing the strong limit of the other summands in (2.4.4), we obtain the strong consistency of $\overline{\sigma}_n^2$

immediately. The same reasoning shows that if the second moment of the stationary distribution is finite (in this case we already know that $\widehat{\boldsymbol{\theta}}^{(n)}$ is a consistent estimator), then the limits of the estimators $\overline{\sigma}_n^2$ and $\widehat{\sigma}_n^2$ are the same almost surely; hence, the strong consistency of $\widehat{\sigma}_n^2$ is established. □

2.4.2 Remark. The CLS estimates are strongly consistent under the null hypothesis only, and the estimation procedure itself supposes that the null hypothesis is valid; however, the calculations can be carried out under the alternative hypothesis as well. Under the alternative hypothesis the weak limit of $\widehat{\boldsymbol{\theta}}_n$ is given in Lemma 2.6.3.

2.5 Construction of the test

We will use a formal analogy between the INAR(p) process and the well-known AR(p) process (Venkataraman, 1982) to obtain analogues of score vector and information quantities as in Gombay (2008). We briefly recall the motivation of the test process as given in T. Szabó (2011b). Due to the martingale central limit theorem,

$$\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} M_k \right)_{t\in[0,1]} \xrightarrow{\mathcal{D}} \left(\sqrt{c}\mathcal{W}_t\right)_{t\in[0,1]},$$

where c is a constant depending on $\boldsymbol{\theta}$ and σ^2, and $(\mathcal{W}_t)_{0\leqslant t\leqslant 1}$ is a standard Brownian motion. Therefore, by a rough approximation

$$(M_1, \dots, M_n) \sim N(0, cE_n),$$

where E_n is the $n \times n$ identity matrix. The approximate likelihood function is

$$\frac{1}{(2\pi c)^{n/2}} \exp\left\{ -\frac{1}{2c^2} \sum_{k=1}^{n} M_k^2 \right\}.$$

We will take the derivative of the log-likelihood function and work with that quantity. The first term will be regarded as constant. This is a simplification because c actually depends on the parameters but taking this into account leads to calculations that are difficult to handle. Also, we will not take into account the constant factor before the sum of the M_k but will rather work with the analogue of the information matrix. Therefore, we consider the following analogue of the loglikelihood function:

$$R_n(\alpha_1, \dots, \alpha_p, \mu) = -2^{-1} \sum_{k=1}^{n} M_k^2.$$

Note that this is the sum that we had to minimize for CLS estimations. The role of the score vector will be played by

$$-\nabla R_k(\widehat{\boldsymbol{\theta}}_n) = \sum_{j=1}^{k} \widehat{M}_j^{(n)} \begin{bmatrix} \boldsymbol{X}_{j-1} \\ 1 \end{bmatrix}.$$

The information matrix $\boldsymbol{I}_n$ is defined by

$$\begin{aligned} \boldsymbol{I}_n &:= \sum_{k=1}^{n} \mathbb{E}\big[\{\nabla R_k(\boldsymbol{\theta}) - \nabla R_{k-1}(\boldsymbol{\theta})\}\{\nabla R_k(\boldsymbol{\theta}) - \nabla R_{k-1}(\boldsymbol{\theta})\}^\top \,|\, \mathcal{F}_{k-1}\big] \\ &= \sum_{k=1}^{n} ((\boldsymbol{\alpha}_*)^\top \boldsymbol{X}_{k-1} + \sigma^2) \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top, \end{aligned} \tag{2.5.1}$$

where $\boldsymbol{\alpha}_*$ comes from (2.3.4). Now we define $\widehat{\boldsymbol{I}}_n$ by replacing in $\boldsymbol{I}_n$ the variance σ^2 and all the parameters in $\boldsymbol{\theta}$ by their CLS estimates. This leads to the $p+1$-dimensional test process $(\widehat{\boldsymbol{\mathcal{M}}}_n(t))_{0\leqslant t\leqslant 1}$ given by

$$\widehat{\boldsymbol{\mathcal{M}}}_n(t) := \widehat{\boldsymbol{I}}_n^{-1/2} \sum_{k=1}^{\lfloor nt\rfloor} \widehat{M}_k^{(n)} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}. \tag{2.5.2}$$

Note that the process $(\widehat{\boldsymbol{\mathcal{M}}}_n(t))_{0\leqslant t\leqslant 1}$ can also be written in the CUSUM form

$$\begin{aligned} \widehat{\boldsymbol{\mathcal{M}}}_n(t) &= \widehat{\boldsymbol{I}}_n^{-1/2} \left(\sum_{k=1}^{\lfloor nt\rfloor} X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} - \boldsymbol{Q}_{\lfloor nt\rfloor} \boldsymbol{Q}_n^{-1} \sum_{k=1}^{n} X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \right) \\ &= \widehat{\boldsymbol{I}}_n^{-1/2} \boldsymbol{Q}_{\lfloor nt\rfloor} \left(\begin{bmatrix} \widehat{\boldsymbol{\alpha}}_{\lfloor nt\rfloor} \\ \widehat{\mu}_{\lfloor nt\rfloor} \end{bmatrix} - \begin{bmatrix} \widehat{\boldsymbol{\alpha}}_n \\ \widehat{\mu}_n \end{bmatrix} \right) = \widehat{\boldsymbol{I}}_n^{-1/2} \boldsymbol{Q}_{\lfloor nt\rfloor} \left(\widehat{\boldsymbol{\theta}}_{\lfloor nt\rfloor} - \widehat{\boldsymbol{\theta}}_n \right). \end{aligned}$$

Under the null hypothesis we have the following result, which allows the construction of various test statistics.

2.5.1 Theorem. *If $(X_k)_{k\geqslant -p+1}$ satisfies H_0 and condition $\mathbf{C}_0$, then*

$$\widehat{\boldsymbol{\mathcal{M}}}_n \xrightarrow{\mathcal{D}} \boldsymbol{\mathcal{B}} \qquad \text{as} \quad n \to \infty,$$

where $(\boldsymbol{\mathcal{B}}(t))_{0\leqslant t\leqslant 1}$ is a $(p+1)$-dimensional standard Brownian bridge, and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution on the Skorokhod space $D([0,1])$.

2.5.2 Remark. This theorem will be sufficient for our tests, and its proof is relatively simple. It is of theoretical interest, however, whether this convergence can be strengthened. It turns out that we can, in fact, prove that there is a series of Brownian bridges that our test process approximates in the strong sense. This will be presented as Theorem A.3.5 later on.

First we show the following application of the martingale central limit theorem.

2.5.3 Theorem. *Let*

$$\boldsymbol{\mathcal{Z}}_n(t) := \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \boldsymbol{Z}_k, \quad t \in [0,1], \qquad \boldsymbol{Z}_k := M_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}, \qquad k = 1, 2, \dots$$

Under the assumptions of Theorem 2.5.1

$$\boldsymbol{\mathcal{Z}}_n \stackrel{\mathcal{D}}{\longrightarrow} \boldsymbol{I}^{1/2} \boldsymbol{\mathcal{W}}, \qquad n \to \infty,$$

where $(\boldsymbol{\mathcal{W}}(t))_{0 \leqslant t \leqslant 1}$ *is a* $(p+1)$*-dimensional standard Wiener process.*

Proof. By (2.3.1) we have

$$\frac{\boldsymbol{I}_n}{n} \stackrel{\text{a.s.}}{\longrightarrow} \boldsymbol{I} := \mathbb{E}\left((\boldsymbol{\alpha}_2^\top \widetilde{\boldsymbol{X}} + \sigma^2) \begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix}^\top \right), \tag{2.5.3}$$

and since $\widehat{\sigma}_n^2$ and $\widehat{\boldsymbol{\theta}}_n$ are strongly consistent estimators, therefore we have

$$n^{-1} \widehat{\boldsymbol{I}}_n \stackrel{\text{a.s.}}{\longrightarrow} \boldsymbol{I} \tag{2.5.4}$$

as well. We will use the martingale central limit theorem for the martingale differences $\frac{1}{\sqrt{n}} \boldsymbol{Z}_k$, $n \in \mathbb{N}$, $k = 1, 2, \dots, n$. To compute the variance function, we write

$$\begin{aligned}
\frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{E}(\boldsymbol{Z}_k \boldsymbol{Z}_k^\top | \mathcal{F}_{k-1}) &= \frac{\lfloor nt \rfloor}{n} \frac{1}{\lfloor nt \rfloor} \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{E}(M_k^2 | \mathcal{F}_{k-1}) \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \\
&= \frac{\lfloor nt \rfloor}{n} \frac{1}{\lfloor nt \rfloor} \sum_{k=1}^{\lfloor nt \rfloor} (\boldsymbol{\alpha}_*^\top \boldsymbol{X}_{k-1} + \sigma^2) \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \\
&\stackrel{\text{a.s.}}{\longrightarrow} t\, \mathbb{E}\left((\boldsymbol{\alpha}_*^\top \widetilde{\boldsymbol{X}} + \sigma^2) \begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix}^\top \right) = t \boldsymbol{I}.
\end{aligned}$$

It remains to check the so-called conditional Lindeberg condition:

$$\begin{aligned}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}\left(\left\|\frac{1}{\sqrt{n}}\boldsymbol{Z}_k\right\|^2\chi_{\{\|n^{-1/2}\boldsymbol{Z}_k\|>\delta\}}\middle|\mathcal{F}_{k-1}\right) &\leqslant \frac{1}{\delta^2}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}\left(\left\|\frac{1}{\sqrt{n}}\boldsymbol{Z}_k\right\|^4\middle|\mathcal{F}_{k-1}\right)\\ &= \frac{1}{\delta^2 n^2}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}\big(M_k^4|\mathcal{F}_{k-1}\big)\big(X_{k-1}^2+\ldots+X_{k-p}^2+1\big)^2\\ &= \frac{1}{\delta^2 n^2}\sum_{k=1}^{\lfloor nt\rfloor}P(\boldsymbol{X}_{k-1}),\end{aligned}$$

where P is a polynomial of degree six, because $\mathbb{E}(M_k^4|\mathcal{F}_{k-1})$ is a second-degree polynomial of $\boldsymbol{X}_{k-1}$ (this is is detailed in section A.2). The sixth moment of the stationary distribution is finite due to the assumptions in $\mathbf{C}_0$, hence (2.3.1) implies

$$\frac{1}{\lfloor nt\rfloor}\sum_{k=1}^{\lfloor nt\rfloor}P(\boldsymbol{X}_{k-1})\xrightarrow{\text{a.s.}}\mathbb{E}(P(\widetilde{\boldsymbol{X}}))<\infty.$$

This means

$$\frac{1}{\delta^2 n^2}\sum_{k=1}^{\lfloor nt\rfloor}P(\boldsymbol{X}_{k-1})\xrightarrow{\text{a.s.}}0,$$

implying Lindeberg's condition. All the conditions of the martingale central limit theorem have been checked; the proof is therefore complete. □

Based on this Theorem, we can use the structure of the estimates to complete the proof of our main result under the null hypothesis.

Proof of Theorem 2.5.1. Let us introduce the notation

$$\widehat{\boldsymbol{Z}}_k^{(n)} := \widehat{M}_k^{(n)}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix},\qquad k=1,2,\ldots$$

First we note that

$$\sum_{k=1}^{\lfloor nt\rfloor}\widehat{\boldsymbol{Z}}_k^{(n)}=\sum_{k=1}^{\lfloor nt\rfloor}\boldsymbol{Z}_k+\sum_{k=1}^{\lfloor nt\rfloor}(\widehat{\boldsymbol{Z}}_k^{(n)}-\boldsymbol{Z}_k)=\sum_{k=1}^{\lfloor nt\rfloor}\boldsymbol{Z}_k+\sum_{k=1}^{\lfloor nt\rfloor}(\widehat{M}_k^{(n)}-M_k)\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}. \tag{2.5.5}$$

Recalling the definitions of M_k and $\widehat{M}_k^{(n)}$,

$$\widehat{M}_k^{(n)}-M_k=\left(X_k-\widehat{\boldsymbol{\theta}}_n^\top\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\right)-\left(X_k-\boldsymbol{\theta}^\top\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\right)=\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_n\right)^\top\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}.$$

Substituting $\widehat{\boldsymbol{\theta}}_n$ from (2.4.1),

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} &= \boldsymbol{Q}_n^{-1}\left(\sum_{k=1}^{n} X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}\right) - \boldsymbol{\theta} \\
&= \boldsymbol{Q}_n^{-1}\left(\sum_{k=1}^{n} X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} - \sum_{k=1}^{n} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^{\top} \boldsymbol{\theta}^{\top}\right) \\
&= \boldsymbol{Q}_n^{-1}\left(\sum_{k=1}^{n} M_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}\right),
\end{aligned} \tag{2.5.6}
$$

hence by (2.5.5),

$$
\widehat{\boldsymbol{I}}_n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \widehat{\boldsymbol{Z}}_k^{(n)} = \sqrt{n}\, \widehat{\boldsymbol{I}}_n^{-1/2} \left(\sum_{k=1}^{\lfloor nt \rfloor} \frac{1}{\sqrt{n}} \boldsymbol{Z}_k - \boldsymbol{Q}_{\lfloor nt \rfloor} \boldsymbol{Q}_n^{-1} \sum_{k=1}^{n} \frac{1}{\sqrt{n}} \boldsymbol{Z}_k \right). \tag{2.5.7}
$$

In the next step we notice that according to (2.3.1),

$$
\boldsymbol{Q}_{\lfloor nt \rfloor} \boldsymbol{Q}_n^{-1} = \frac{\lfloor nt \rfloor}{n} \left(\frac{1}{\lfloor nt \rfloor} \boldsymbol{Q}_{\lfloor nt \rfloor} \right) \left(\frac{1}{n} \boldsymbol{Q}_n \right)^{-1} \xrightarrow{\text{a.s.}} t \widetilde{\boldsymbol{Q}}_{\mathrm{H}_0} \widetilde{\boldsymbol{Q}}_{\mathrm{H}_0}^{-1} = t \boldsymbol{E}_{p+1} \qquad \forall t \in [0,1],
$$

where $\boldsymbol{E}_{p+1}$ is the $p+1$-dimensional identity matrix and

$$
\widetilde{\boldsymbol{Q}}_{\mathrm{H}_0} := \mathbb{E}\left(\begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{X}} \\ 1 \end{bmatrix}^{\top} \right).
$$

Now we apply (2.5.7), Theorem 2.5.3, and (2.5.4) to conclude that

$$
\left(\widehat{\boldsymbol{I}}_n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \widehat{\boldsymbol{Z}}_k^{(n)} \right)_{t \in [0,1]} \xrightarrow{\mathcal{D}} (\boldsymbol{\mathcal{W}}(t) - t\boldsymbol{\mathcal{W}}(1))_{t \in [0,1]}.
$$

This completes our proof. □

2.5.1 Testing procedures

By the continuous mapping theorem we obtain the following corollary of Theorem 2.5.1.

2.5.4 Corollary. *Under the assumptions of Theorem 2.5.1 we have*

$$\sup_{0\leqslant t\leqslant 1} \widehat{\mathcal{M}}_n^{(i)}(t) \stackrel{\mathcal{D}}{\longrightarrow} \sup_{0\leqslant t\leqslant 1} B(t), \tag{2.5.8}$$

$$\inf_{0\leqslant t\leqslant 1} \widehat{\mathcal{M}}_n^{(i)}(t) \stackrel{\mathcal{D}}{\longrightarrow} \inf_{0\leqslant t\leqslant 1} B(t), \tag{2.5.9}$$

$$\sup_{0\leqslant t\leqslant 1} |\widehat{\mathcal{M}}_n^{(i)}(t)| \stackrel{\mathcal{D}}{\longrightarrow} \sup_{0\leqslant t\leqslant 1} |B(t)| \tag{2.5.10}$$

as $n \to \infty$, *where* $(\widehat{\mathcal{M}}_n^{(i)}(t))_{0\leqslant t\leqslant 1}$, $i = 1, \dots, p+1$, *denotes the components of* $(\widehat{\boldsymbol{\mathcal{M}}}_n(t))_{0\leqslant t\leqslant 1}$, *and* $(B(t))_{0\leqslant t\leqslant 1}$ *is a standard Brownian bridge.*

Since $(\boldsymbol{\mathcal{B}}(t))_{0\leqslant t\leqslant 1}$ in Theorem 2.5.1 has independent components, we need to define the tests component-wise only. For simultaneous test-for-change in d parameters, to have an overall level of significance α, we use $\alpha^* := 1 - (1-\alpha)^{1/d}$ for each component. We can test for change in a single component, $\theta^{(i)}$, $i \in PV$ (with $\theta^{(i)} = \alpha_i$ for $i = 1, 2, \dots, p$ and $\theta^{(p+1)} = \mu$, according to the definition of $\boldsymbol{\theta}$) in the following way:

Two different tests can be constructed:

Test 1 (one-sided): If

$$\sup_{0\leqslant t\leqslant 1} \widehat{\mathcal{M}}_n^{(i)}(t) \geqslant C_1(\alpha^*) \quad \text{or} \quad \inf_{0\leqslant t\leqslant 1} \widehat{\mathcal{M}}_n^{(i)}(t) \leqslant -C_1(\alpha^*),$$

then we conclude that there was a downward or upward change in parameter $\theta^{(i)}$ (respectively) along the sequence $X_0, X_1, \dots, X_n$.

Test 2 (two-sided): If

$$\sup_{0\leqslant t\leqslant 1} |\widehat{\mathcal{M}}_n^{(i)}(t)| \geqslant C_2(\alpha^*),$$

then we conclude that there was a change in parameter $\theta^{(i)}$ along the sequence $X_0, X_1, \dots, X_n$.

Critical values are obtained from the limit distributions in Corollary 2.5.4, namely, from the identities

$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant 1} B(t) \geqslant x\right) = \mathrm{e}^{-2x^2}, \quad x \geqslant 0,$$

$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant 1} |B(t)| \geqslant x\right) = 2\sum_{k=1}^{\infty} (-1)^{k+1} \mathrm{e}^{-2k^2x^2}, \quad x \geqslant 0,$$

respectively, where $(B(t))_{0\leqslant t\leqslant 1}$ is a Brownian bridge.

2.6 The process under the alternative hypothesis

While in Theorem 2.5.1 we were able to consider longer and longer samples taken from the same process, this approach has to be modified for the alternative hypothesis. More precisely, we have to consider a series of time-inhomogeneous INAR(p) processes, where the n-th one has a point of change at $\lfloor n\rho \rfloor$ (we will suppress this in the notation for simplicity). Now, the parts of these processes before the change (i.e., $(X_i)_{i=1}^{\lfloor n\rho \rfloor}$) can be handled as a sample taken from an infinite INAR(p) process (at least in distribution), but this is not true for the second part (i.e., $(X_i)_{i \geqslant \lfloor n\rho \rfloor + 1}$), because the initial distribution of this process depends on n. Therefore, for a rigorous analysis we need to refine the results of 2.3. We will impose some additional conditions on the parameters of the process, which are summarized below.

2.6.1 Definition. *We will say that an INAR(p) process* $(X_k)_{k \geqslant -p+1}$ *satisfies* $\mathbf{C}_{\mathrm{A}}$ *if* $\tau = \lfloor n\rho \rfloor$ *for some* $\rho \in (0,1)$*, both* $(X_k)_{-p+1 \leqslant k \leqslant \tau}$ *and* $(X_k)_{k \geqslant \tau+1}$ *satisfy condition* $\mathbf{C}_0$*, and the parameter vectors for the processes* $(X_k)_{-p+1 \leqslant k \leqslant \tau}$ *and* $(X_k)_{k \geqslant \tau+1}$ *are*

$$\boldsymbol{\theta}' := \begin{bmatrix} \boldsymbol{\alpha}' \\ \mu' \end{bmatrix}, \quad \textit{and} \quad \boldsymbol{\theta}'' := \begin{bmatrix} \boldsymbol{\alpha}'' \\ \mu'' \end{bmatrix},$$

respectively. In this case, $\widetilde{\boldsymbol{X}}'$ *and* $\widetilde{\boldsymbol{X}}''$ *will denote variables with the unique stationary distributions of the two halves of the process, respectively. We will use the following notations:*

$$\boldsymbol{Q}' := \mathbb{E}\left(\begin{bmatrix} \widetilde{\boldsymbol{X}}' \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{X}}' \\ 1 \end{bmatrix}^\top \right), \qquad \boldsymbol{Q}'' := \mathbb{E}\left(\begin{bmatrix} \widetilde{\boldsymbol{X}}'' \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{X}}'' \\ 1 \end{bmatrix}^\top \right), \qquad \widetilde{\boldsymbol{Q}} := \rho \boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}''.$$

2.6.1 Ergodicity

The equivalent of (2.3.1) is the following.

2.6.2 Lemma. *Under* $\mathbf{C}_{\mathrm{A}}$ *we have*

$$\frac{1}{n - \lfloor n\rho \rfloor} \sum_{k=\lfloor n\rho \rfloor + 1}^{n} g(\boldsymbol{X}_k) \xrightarrow{\mathbb{P}} \mathbb{E}(g(\widetilde{\boldsymbol{X}}'')), \tag{2.6.1}$$

whenever $g : \mathbb{Z}_+^p \to \mathbb{R}$ *with* $\mathbb{E}(|g(\widetilde{\boldsymbol{X}}'')|) < \infty$.

Proof. For an arbitrary $\varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{1}{n-\lfloor n\rho\rfloor}\sum_{k=\lfloor n\rho\rfloor+1}^{n} g(\boldsymbol{X}_k) - \mathbb{E}(g(\widetilde{\boldsymbol{X}}''))\right| > \varepsilon\right)$$
$$= \sum_{x\in\mathbb{Z}_+^p} \mathbb{P}\left(\left|\frac{1}{n-\lfloor n\rho\rfloor}\sum_{k=\lfloor n\rho\rfloor+1}^{n} g(\boldsymbol{X}_k) - \mathbb{E}(g(\widetilde{\boldsymbol{X}}''))\right| > \varepsilon \,\middle|\, \boldsymbol{X}_{\lfloor n\rho\rfloor} = x\right) \mathbb{P}(\boldsymbol{X}_{\lfloor n\rho\rfloor} = x),$$

and

$$\mathbb{P}\left(\left|\frac{1}{n-\lfloor n\rho\rfloor}\sum_{k=\lfloor n\rho\rfloor+1}^{n} g(\boldsymbol{X}_k) - \mathbb{E}(g(\widetilde{\boldsymbol{X}}''))\right| > \varepsilon \,\middle|\, \boldsymbol{X}_{\lfloor n\rho\rfloor} = x\right) \to 0$$

by the ergodic theorem for each $x \in \mathbb{Z}_+^p$, additionally

$$\mathbb{P}(\boldsymbol{X}_{\lfloor n\rho\rfloor} = x) \leqslant |\,\mathbb{P}(\boldsymbol{X}_{\lfloor n\rho\rfloor} = x) - \mathbb{P}(\widetilde{\boldsymbol{X}} = x)| + \mathbb{P}(\widetilde{\boldsymbol{X}} = x),$$

and one can use (2.3.3). □

We will also apply that for all $\varepsilon > 0$ there exists ν such that

$$\|\,\mathbb{E}(\boldsymbol{X}_{\lfloor n\rho\rfloor+k}) - \mathbb{E}(\widetilde{\boldsymbol{X}}'')\| < \varepsilon \qquad \text{for all } n \geqslant \nu \text{ and all } k \geqslant \nu. \tag{2.6.2}$$

For this, first observe that as a consequence of (2.3.5), there exists $\pi'' \in (0,1)$ such that

$$\|\,\mathbb{E}(\boldsymbol{X}_{\lfloor n\rho\rfloor+k}) - \mathbb{E}(\widetilde{\boldsymbol{X}}'')\| \leqslant (\pi'')^k \|\,\mathbb{E}(\boldsymbol{X}_{\lfloor n\rho\rfloor}) - \mathbb{E}(\widetilde{\boldsymbol{X}}'')\| \quad k \in \mathbb{N}.$$

Next, for all $\eta > 0$, choose ν_1 and ν_2 such that $(\pi'')^k < \eta$ for all $k \geqslant \nu_1$ and $\|\,\mathbb{E}(\boldsymbol{X}_{\lfloor n\rho\rfloor}) - \mathbb{E}(\widetilde{\boldsymbol{X}}')\| < \eta$ for all $n \geqslant \nu_2$. Hence

$$\begin{aligned}\|\,\mathbb{E}(\boldsymbol{X}_{\lfloor n\rho\rfloor+k}) - \mathbb{E}(\widetilde{\boldsymbol{X}}'')\| &\leqslant \eta\big(\|\,\mathbb{E}(\boldsymbol{X}_{\lfloor n\rho\rfloor}) - \mathbb{E}(\widetilde{\boldsymbol{X}}')\| + \|\,\mathbb{E}(\widetilde{\boldsymbol{X}}') - \mathbb{E}(\widetilde{\boldsymbol{X}}'')\|\big)\\ &\leqslant \eta^2 + \eta\|\,\mathbb{E}(\widetilde{\boldsymbol{X}}') - \mathbb{E}(\widetilde{\boldsymbol{X}}'')\|.\end{aligned}$$

For the behavior of the CLS estimates under $\mathbf{C}_{\mathrm{A}}$, we have the following result.

2.6.3 Lemma. *Under* $\mathbf{C}_{\mathrm{A}}$ *we have*

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathbb{P}} \widetilde{\boldsymbol{\theta}} := \widetilde{\boldsymbol{Q}}^{-1}\left(\rho\boldsymbol{Q}'\boldsymbol{\theta}' + (1-\rho)\boldsymbol{Q}''\boldsymbol{\theta}''\right)$$

and

$$\frac{\widehat{\boldsymbol{I}}_n}{n} \xrightarrow{\mathbb{P}} \widetilde{\boldsymbol{I}} := \mathbb{E}\left(\rho(\boldsymbol{\alpha}_*'^{\top}\widetilde{\boldsymbol{X}} + \sigma^2)\begin{bmatrix}\widetilde{\boldsymbol{X}}'\\ 1\end{bmatrix}\begin{bmatrix}\widetilde{\boldsymbol{X}}'\\ 1\end{bmatrix}^{\top} + (1-\rho)(\boldsymbol{\alpha}_*''^{\top}\widetilde{\boldsymbol{X}} + \sigma^2)\begin{bmatrix}\widetilde{\boldsymbol{X}}''\\ 1\end{bmatrix}\begin{bmatrix}\widetilde{\boldsymbol{X}}''\\ 1\end{bmatrix}^{\top}\right).$$

Proof. By (2.3.1) and (2.6.1) we obtain

$$\frac{1}{n}\boldsymbol{Q}_n = \frac{1}{n}\boldsymbol{Q}_{\lfloor n\rho \rfloor} + \frac{1}{n}\sum_{k=\lfloor n\rho\rfloor+1}^{n} \begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^{\top} \xrightarrow{\mathbb{P}} \rho\boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}'' = \widetilde{\boldsymbol{Q}},$$

as $n \to \infty$. Moreover, exactly as in (2.4.8),

$$\frac{1}{n}\sum_{k=1}^{\lfloor n\rho\rfloor} X_k \begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix} \xrightarrow{\mathbb{P}} \rho\boldsymbol{Q}' \begin{bmatrix}\boldsymbol{\alpha}\\ \mu'\end{bmatrix} \tag{2.6.3}$$

as $n \to \infty$. In a similar way, using (2.6.1)

$$\frac{1}{n}\sum_{k=\lfloor n\rho\rfloor+1}^{n} X_k \begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix} \xrightarrow{\mathbb{P}} (1-\rho)\boldsymbol{Q}'' \begin{bmatrix}\boldsymbol{\alpha}\\ \mu''\end{bmatrix}.$$

The second statement of the lemma can be proved in the same way by an analogy with (2.5.3). □

2.7 Consistency of the test

The following theorem, the analogue of Theorem 3.1 in Hušková et al. (2007), describes the behavior of the maximum of the test process if a change occurs in the mean of the innovation. An immediate consequence of the theorem is that the test statistic tends to infinity stochastically as $n \to \infty$, which suffices for the weak consistency of the proposed test. For further discussion of this result, see Remark 4.5.4, which applies here equally.

2.7.1 Theorem. *Suppose that $\boldsymbol{C}_A$ holds. For $i = 1, 2, \ldots, p+1$, let us define*

$$\psi_i := \mathbf{1}_i^{\top}\widetilde{\boldsymbol{I}}^{-1/2}((\rho\boldsymbol{Q}')^{-1} + ((1-\rho)\boldsymbol{Q}'')^{-1})^{-1}(\boldsymbol{\theta}' - \boldsymbol{\theta}'').$$

If $\psi_i > 0$ then for the i-th component of the test process,

$$\sup_{0\leqslant t\leqslant 1} \widehat{\boldsymbol{\mathcal{M}}}_n(t)^{(i)} = n^{1/2}\psi_i + \mathrm{o}_{\mathbb{P}}(n^{1/2}),$$

and conversely, if $\psi_i < 0$ then

$$\inf_{0\leqslant t\leqslant 1} \widehat{\boldsymbol{\mathcal{M}}}_n(t)^{(i)} = n^{1/2}\psi_i + \mathrm{o}_{\mathbb{P}}(n^{1/2}).$$

Proof. We will only prove for $i = 1$ and $\psi_1 > 0$, the other cases are completely analogous.

We will use the following notations:

$$M_k' := X_k - \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \boldsymbol{\theta}, \qquad \boldsymbol{Z}_k' := M_k' \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} = X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} - \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \boldsymbol{\theta}',$$

and similarly for M_k'' and $\boldsymbol{Z}_k''$. The proof will be given for the process before $\lfloor n\rho \rfloor$ in detail. The analysis of the process after $\lfloor n\rho \rfloor$ can be handled analogously. In the proof we will rely repeatedly on ideas from Hušková et al. (2007). The task is essentially to determine the weak limit of the supremum of

$$\begin{aligned} n^{-1/2}\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t) &= n^{-1}\mathbf{1}_1 \widetilde{\boldsymbol{I}}^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \widehat{M}_k^{(n)} \\ &\quad + n^{-1}\mathbf{1}_1 \left(\left(\frac{\widehat{\boldsymbol{I}}_n}{n} \right)^{-1/2} - \widetilde{\boldsymbol{I}}^{-1/2} \right) \sum_{k=1}^{\lfloor nt \rfloor} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \widehat{M}_k^{(n)}. \end{aligned} \tag{2.7.1}$$

For the first term of (2.7.1) we apply the following decomposition for $k < \lfloor n\rho \rfloor$:

$$\begin{aligned} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \widehat{M}_k^{(n)} &= \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \left(X_k - \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \widehat{\boldsymbol{\theta}}_n \right) \\ &= \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \left(X_k - \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \boldsymbol{\theta}' \right) + \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \left(\boldsymbol{\theta}' - \widehat{\boldsymbol{\theta}}_n \right) \\ &= \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} M_k' + \mathbb{E} \left(\begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \right) (\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}}) \\ &\quad + \left(\mathbb{E} \left(\begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \right) - \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \right) (\boldsymbol{\theta}' - \widehat{\boldsymbol{\theta}}_n) \\ &\quad + \mathbb{E} \left(\begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \right) (\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_n), \end{aligned} \tag{2.7.2}$$

and similarly if we replace M_k' and $\boldsymbol{\theta}'$ with M_k'' and $\boldsymbol{\theta}''$, respectively.

Based on (2.7.2),

(2.7.3)

$$
\begin{aligned}
S(0,\lfloor nt\rfloor) &:= n^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sum_{k=1}^{\lfloor nt\rfloor}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\widehat{M}_k^{(n)} \\
&= n^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sum_{k=1}^{\lfloor nt\rfloor}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}M_k' \\
&\quad + n^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right)(\boldsymbol{\theta}'-\widetilde{\boldsymbol{\theta}}) \\
&\quad + n^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sum_{k=1}^{\lfloor nt\rfloor}\left(\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right)-\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right)(\boldsymbol{\theta}'-\widehat{\boldsymbol{\theta}}_n) \\
&\quad + n^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right)(\widetilde{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_n) \\
&=: S_1(0,\lfloor nt\rfloor,\boldsymbol{\theta}')+S_2(0,\lfloor nt\rfloor,\boldsymbol{\theta}')+S_3(0,\lfloor nt\rfloor,\boldsymbol{\theta}')+S_4(0,\lfloor nt\rfloor),
\end{aligned}
$$

and similarly,

$$
S(0,\lfloor nt\rfloor)=S_1(0,\lfloor nt\rfloor,\boldsymbol{\theta}'')+S_2(0,\lfloor nt\rfloor,\boldsymbol{\theta}'')+S_3(0,\lfloor nt\rfloor,\boldsymbol{\theta}'')+S_4(0,\lfloor nt\rfloor).
$$

Introducing now $S_i(a,b,\boldsymbol{\theta}) := S_i(0,b,\boldsymbol{\theta})-S_i(0,a,\boldsymbol{\theta}),\ i=1,2,\ldots,p+1$, the quantity that interests us is

(2.7.4)

$$
\begin{aligned}
\left|\sup_{t\in[0,1]}S(0,\lfloor nt\rfloor)-\psi_1\right| \leqslant{}& \sup_{t\in[0,1]}\left|S_1(0,\lfloor n\rho\rfloor\wedge\lfloor nt\rfloor,\boldsymbol{\theta}')+S_1(\lfloor n\rho\rfloor\wedge\lfloor nt\rfloor,\lfloor nt\rfloor,\boldsymbol{\theta}'')\right| \\
&+\left|\sup_{t\in[0,1]}\left(S_2(0,\lfloor n\rho\rfloor\wedge\lfloor nt\rfloor,\boldsymbol{\theta}')+S_2(\lfloor n\rho\rfloor\wedge\lfloor nt\rfloor,\lfloor nt\rfloor,\boldsymbol{\theta}'')\right)-\psi_1\right| \\
&+\sup_{t\in[0,1]}\left|S_3(0,\lfloor n\rho\rfloor\wedge\lfloor nt\rfloor,\boldsymbol{\theta}')+S_3(\lfloor n\rho\rfloor\wedge\lfloor nt\rfloor,\lfloor nt\rfloor,\boldsymbol{\theta}'')\right| \\
&+\sup_{t\in[0,1]}\left|S_4(0,\lfloor nt\rfloor)\right|.
\end{aligned}
$$

The first, third and fourth terms in (2.7.4) are all $o_{\mathbb{P}}(1)$ according to Lemmas 2.9.3, 2.9.2 and 2.9.4, respectively. All that remains is the second term. Let us notice here that $\boldsymbol{Q}'$ and $\boldsymbol{Q}''$ are both symmetric, which we will exploit repeatedly. It is evident from the definition of $\widetilde{\boldsymbol{\theta}}$

(see Lemma 2.6.3) that

$$\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}' - (1-\rho)(\rho \boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}'')^{-1}\boldsymbol{Q}''(\boldsymbol{\theta}' - \boldsymbol{\theta}''),$$

and so

$$\begin{aligned}\mathbf{1}_1 \widetilde{\boldsymbol{I}}^{-1/2}\, \mathbb{E}\left(\begin{bmatrix}\widetilde{\boldsymbol{X}}' \\ 1\end{bmatrix}\begin{bmatrix}\widetilde{\boldsymbol{X}}' \\ 1\end{bmatrix}^\top\right)(\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}}) &= \mathbf{1}_1 \widetilde{\boldsymbol{I}}^{-1/2} \boldsymbol{Q}'(1-\rho)(\rho \boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}'')^{-1}\boldsymbol{Q}''(\boldsymbol{\theta}' - \boldsymbol{\theta}'') \\ &= \frac{\psi_1}{\rho},\end{aligned}$$

and similarly,

$$\begin{aligned}\mathbf{1}_1 \widetilde{\boldsymbol{I}}^{-1/2}\, \mathbb{E}\left(\begin{bmatrix}\widetilde{\boldsymbol{X}}'' \\ 1\end{bmatrix}\begin{bmatrix}\widetilde{\boldsymbol{X}}'' \\ 1\end{bmatrix}^\top\right)(\boldsymbol{\theta}'' - \widetilde{\boldsymbol{\theta}}) &= \mathbf{1}_1 \widetilde{\boldsymbol{I}}^{-1/2} \boldsymbol{Q}''(1-\rho)(\rho \boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}'')^{-1}\boldsymbol{Q}''(\boldsymbol{\theta}' - \boldsymbol{\theta}'') \\ &= \frac{\psi_1}{1-\rho},\end{aligned}$$

since, by an easy calculation,

$$\left((\rho \boldsymbol{Q}')^{-1} + ((1-\rho)\boldsymbol{Q}'')^{-1}\right)^{-1} = \rho(1-\rho)\left(\boldsymbol{Q}'\left(\rho \boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}''\right)^{-1}\boldsymbol{Q}''\right).$$

Now we can write

$$\begin{aligned}&\left|\sup_{t\in[0,1]}\left(S_2(0, \lfloor n\rho\rfloor \wedge \lfloor nt\rfloor, \boldsymbol{\theta}') + S_2(\lfloor n\rho\rfloor \wedge \lfloor nt\rfloor, \lfloor nt\rfloor, \boldsymbol{\theta}'') - \psi_1\right)\right| \\ &\leqslant \left|\sup_{t\in[0,1]}\left(\frac{\lfloor n\rho\rfloor \wedge \lfloor nt\rfloor}{n\rho} - \frac{(\lfloor nt\rfloor - \lfloor n\rho\rfloor)^+}{n(1-\rho)} - 1\right)\psi_1\right| \\ &\quad + T^{-1}\mathbf{1}_1 \widetilde{\boldsymbol{I}}^{-1/2} \sup_{t\in[0,1]}\left|\int_0^{(\rho T)\wedge t}\begin{bmatrix}0 & \mathbb{E}(Y'_\infty) - \mathbb{E}(Y_u) \\ \mathbb{E}(Y'_\infty) - \mathbb{E}(Y_u) & \mathbb{E}(Y_u^2) - \mathbb{E}(Y'^2_\infty)\end{bmatrix}^\top(\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}})\,\mathrm{d}u\right| \\ &\quad + T^{-1}\mathbf{1}_1 \widetilde{\boldsymbol{I}}^{-1/2} \sup_{t\in[0,1]}\left|\int_0^{(\rho T)\wedge t}\begin{bmatrix}0 & \mathbb{E}(Y'_\infty) - \mathbb{E}(Y_u) \\ \mathbb{E}(Y'_\infty) - \mathbb{E}(Y_u) & \mathbb{E}(Y_u^2) - \mathbb{E}(Y'^2_\infty)\end{bmatrix}^\top(\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}})\,\mathrm{d}u\right|.\end{aligned}$$

The second and third terms converge to 0 by Lemma 2.3.1. We conclude the proof for the first term in (2.7.1) by noting that the supremum of the first term is clearly attained at $t = \rho$ and is

$$1 - \frac{\lfloor n\rho\rfloor}{n\rho} \to 0.$$

All that remains is showing that the second term in (2.7.1) is $\mathrm{o}_{\mathbb{P}}(1)$. To see this, consider, with the L_1-norm $\|\cdot\|$, and its induced matrix norm $\|\cdot\|_*$,

$$\begin{aligned}\sup_{t\in[0,1]} n^{-1}\mathbf{1}_1 \left(\left(\frac{\widehat{\boldsymbol{I}}_n}{n}\right)^{-1/2} - \widetilde{\boldsymbol{I}}^{-1/2}\right) \sum_{k=1}^{\lfloor nt\rfloor} \begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix} \widehat{M}_k^{(n)} \\ \leqslant \left\|\left(\frac{\widehat{\boldsymbol{I}}_n}{n}\right)^{-1/2} - \widetilde{\boldsymbol{I}}^{-1/2}\right\|_* \left\|\widetilde{\boldsymbol{I}}^{1/2}\right\|_* \sup_{t\in[0,1]} \left\| n^{-1}\widetilde{\boldsymbol{I}}^{-1/2} \sum_{k=1}^{\lfloor nt\rfloor} \begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix} \widehat{M}_k^{(n)} \right\|,\end{aligned} \tag{2.7.5}$$

which is clearly $\mathrm{o}_{\mathbb{P}}(1)$ since the first factor is $\mathrm{o}_{\mathbb{P}}(1)$ (note that $\widetilde{\boldsymbol{I}}$ is invertible, hence we can use Lemma 2.6.3 and the continuous mapping theorem), the second factor is finite, and the third has just been shown to be $K + \mathrm{o}_{\mathbb{P}}(1)$ for some constant K. □

2.8 Estimation of the change point

Based on the score vector analogy, if there is a change in the i-th parameter, the estimator of τ is

$$\widehat{\tau}_n := n \inf\left\{t \in (0,1) : \widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t) = \sup_{0<t<1} \widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\right\} \tag{2.8.1}$$

for the downward one-sided test,

$$\widehat{\tau}_n := n \inf\left\{t \in (0,1) : \widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t) = \inf_{0<t<1} \widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\right\} \tag{2.8.2}$$

for the upward one-sided test, and

$$\widehat{\tau}_n := n \inf\left\{t \in (0,1) : \left|\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\right| = \sup_{0<t<1} \left|\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\right|\right\} \tag{2.8.3}$$

for the two-sided test.

2.8.1 Theorem. *If $\mathbf{C}_A$ holds, then we have*

$$\widehat{\tau}_n - \lfloor n\rho\rfloor = \mathrm{O}_{\mathbb{P}}(1) \qquad \textit{as} \;\; n \to \infty.$$

Consequently, if we define $\widehat{\rho}_n := \frac{\widehat{\tau}_n}{n}$, then $\widehat{\rho}_n - \rho = \mathrm{O}_{\mathbb{P}}(n^{-1})$.

2.8.2 Remark. This result is slightly stronger than the similar Proposition 3.1 in Hušková et al. (2007). Similar results are valid for change in a location parameter (see Csörgő and Horváth, 1997), and in these cases the limit distribution is nondegenerate. Therefore we can conjecture that Theorem 2.8.1 cannot be improved upon in terms of convergence rate.

Proof. The statement can be written in the form

$$\lim_{K\to\infty}\sup_{n\in\mathbb{N}}\mathbb{P}(|\widehat{\tau}_n-\lfloor n\rho\rfloor|\geqslant K)=0,$$

which is equivalent to

$$\lim_{K\to\infty}\limsup_{n\to\infty}\mathbb{P}(|\widehat{\tau}_n-\lfloor n\rho\rfloor|\geqslant K)=0.$$

Hence to prove the statement it is enough to show that

$$\lim_{K\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\max_{\rho-\frac{K}{n}<t<\rho+\frac{K}{n}}\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\leqslant\max_{0<t\leqslant\rho-\frac{K}{n}}\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\right)=0,\tag{2.8.4}$$

$$\lim_{K\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\max_{\rho-\frac{K}{n}<t<\rho+\frac{K}{n}}\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\leqslant\max_{\rho+\frac{K}{n}<t<1}\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\right)=0.\tag{2.8.5}$$

For (2.8.4) we consider with a constant K, $K<\lfloor n\rho\rfloor$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\rho-\frac{K}{n}<t<\rho+\frac{K}{n}}\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\leqslant\sup_{0<t\leqslant\rho-\frac{K}{n}}\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\right)\\
&\qquad\leqslant\mathbb{P}\left(\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(\rho)\leqslant\sup_{0<t\leqslant\rho-\frac{K}{n}}\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\right)=\mathbb{P}\left(\inf_{0<t\leqslant\rho-\frac{K}{n}}\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(\rho)-\widehat{\boldsymbol{\mathcal{M}}}_n^{(i)}(t)\leqslant 0\right)\\
&\qquad=\mathbb{P}\left(\inf_{K\leqslant\ell\leqslant\lfloor n\rho\rfloor-1}\mathbf{1}_i\widehat{\boldsymbol{I}}_n^{-1/2}\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\begin{bmatrix}\boldsymbol{X}_{k-1}\\1\end{bmatrix}\widehat{M}_j^{(n)}\leqslant 0\right)\\
&\qquad=\mathbb{P}\left(\inf_{K\leqslant\ell\leqslant\lfloor n\rho\rfloor-1}n^{1/2}\ell^{-1}\mathbf{1}_i\widehat{\boldsymbol{I}}_n^{-1/2}\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\begin{bmatrix}\boldsymbol{X}_{k-1}\\1\end{bmatrix}\widehat{M}_j^{(n)}\leqslant 0\right).
\end{aligned}$$

As in Theorem 2.7.1, we only prove for $i=1$. For any $K\leqslant\ell\leqslant\lfloor n\rho\rfloor$ the expression

$$n^{1/2}\ell^{-1}\mathbf{1}_1\widehat{\boldsymbol{I}}_n^{-1/2}\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\begin{bmatrix}\boldsymbol{X}_{k-1}\\1\end{bmatrix}\widehat{M}_j^{(n)}\tag{2.8.6}$$

can be decomposed into five terms in a similar way to (2.7.1) and (2.7.2). Now, (2.8.6) can only be negative in two cases: either the dominant term in the decomposition is less than or equal to $\frac{\psi_i}{2}$, or it is greater—in which case one of the other four terms has to be less than $-\frac{\psi_i}{8}$ (for the definition of ψ_i, see Theorem 2.7.1).

To make what we said concrete,

$$
\begin{aligned}
&\mathbb{P}\left(\inf_{K\leqslant\ell\leqslant\lfloor n\rho\rfloor-1} n^{1/2}\ell^{-1}\mathbf{1}_i\widehat{\boldsymbol{I}}_n^{-1/2}\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\widehat{M}_j^{(n)}\leqslant 0\right)\\
&\leqslant \mathbb{P}\left(\min_{K\leqslant\ell\leqslant\lfloor n\rho\rfloor-1}\ell^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right)(\boldsymbol{\theta}'-\widetilde{\boldsymbol{\theta}})\leqslant\frac{\psi_1}{2}\right)\\
&\quad+\mathbb{P}\left(\max_{K\leqslant\ell\leqslant\lfloor n\rho\rfloor-1}\left|\ell^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}M_k'\right|\geqslant\frac{\psi_1}{8}\right)\\
&\quad+\mathbb{P}\left(\frac{\psi_1}{8}\leqslant\max_{K\leqslant\ell\leqslant\lfloor n\rho\rfloor-1}\left|\ell^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\left(\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right)\right.\right.\right.\\
&\qquad\qquad\left.\left.\left.-\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right)(\boldsymbol{\theta}'-\widehat{\boldsymbol{\theta}}_n)\right|\right)\\
&\quad+\mathbb{P}\left(\max_{K\leqslant\ell\leqslant\lfloor n\rho\rfloor-1}\left|\ell^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right)(\widetilde{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_n)\right|\geqslant\frac{\psi_1}{8}\right)\\
&\quad+\mathbb{P}\left(\sup_{K\leqslant\ell\leqslant\lfloor n\rho\rfloor-1}\left|\ell^{-1}\mathbf{1}_1\left(\left(\frac{\widehat{\boldsymbol{I}}_n}{n}\right)^{-1/2}-\widetilde{\boldsymbol{I}}^{-1/2}\right)\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\widehat{M}_j^{(n)}\right|\geqslant\frac{\psi_1}{8}\right).
\end{aligned}
$$

As a consequence of (2.3.2) the first term can be shown to converge to zero for any K as $n\to\infty$. This is a rather elementary exercise in calculus and it is proven in Lemma 2.9.5. Because of (2.3.2) and Lemma 2.9.4, the fourth term also converges to zero for all K as $n\to\infty$. Indeed, $(\widetilde{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_n)\xrightarrow{\mathbb{P}}0$ and

$$
\max_{K\leqslant\ell\leqslant\lfloor n\rho\rfloor}\ell^{-1}\sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor}\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right)\leqslant\max_{1\leqslant k\leqslant\lfloor n\rho\rfloor}\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top\right),
$$

and due to (2.3.2) the right hand side is bounded as $n\to\infty$. The convergence of the second and third terms is the statement of Lemma 2.9.7. The fifth term can be handled in the same way as in (2.7.5). To prove (2.8.5) the proof is analogous with one exception: in place of Lemma 2.9.5 we need Lemma 2.9.6. □

2.9 Lemmas for Theorems 2.7.1 and 2.8.1

The lemmas collected here are crucial to the proofs of our main theorems, but their proofs are somewhat tedious, hence they have been collected here together for the interested reader.

Lemmas 2.9.2, 2.9.3 and 2.9.7 each contain two similar statements – one about $\boldsymbol{X}_k$ and another about $\boldsymbol{X}_k^{\otimes 2}$. In both cases, we will only prove the first statement – the second one can always be proved in the same manner by using (ii) from Lemma 2.3.3 instead of (i) from the same Lemma. First we recall a Hájek–Rényi type result that will be critical not only here, but also in the continuous case.

2.9.1 Lemma. (Kokoszka and Leipus, 1998, Theorem 3.1) *Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random variables with finite second moments, and let $(c_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative constants. Then, for any $a > 0$,*

$$\begin{aligned} a^2\,\mathbb{P}\left(\max_{1\leqslant k\leqslant n} c_k \left|\sum_{j=1}^k Y_j\right| > a\right) &\leqslant \sum_{k=1}^{n-1} |c_{k+1}^2 - c_k^2| \sum_{i,j=1}^k \mathbb{E}(Y_iY_j) \\ &\quad + 2\sum_{k=1}^{n-1} c_{k+1}^2 \left(\mathbb{E}\left(Y_{k+1}^2\right) \sum_{i,j=1}^k \mathbb{E}\left(Y_iY_j\right)\right)^{1/2} \\ &\quad + 2\sum_{k=0}^{n-1} c_{k+1}^2\,\mathbb{E}(Y_{k+1}^2). \end{aligned}$$

2.9.2 Lemma. *For a time-homogeneous INAR(p) process satisfying condition $\mathbf{C}_0$ and any $\gamma < \frac{1}{4}$ we have*

$$\max_{1\leqslant k\leqslant n} k^{\gamma-1}\left\|\sum_{i=1}^k (\boldsymbol{X}_{k-1} - \mathbb{E}(\boldsymbol{X}_{k-1}))\right\| = \mathrm{O}_{\mathbb{P}}(1)$$

and

$$\max_{1\leqslant k\leqslant n} k^{\gamma-1}\left\|\sum_{i=1}^k \left(\boldsymbol{X}_{k-1}^{\otimes 2} - \mathbb{E}\left(\boldsymbol{X}_{k-1}^{\otimes 2}\right)\right)\right\| = \mathrm{O}_{\mathbb{P}}(1).$$

Proof. We will follow the proof of Lemma 4.2 in Hušková et al. (2007) and apply Lemma 2.9.1 with $c_k = k^{\gamma-1}$ and $Y_{i,q} = X_{i-1-q} - \mathbb{E}(X_{i-1-q})$ for $0 \leqslant q \leqslant p-1$ to show that the result holds for each component of the vectors. This implies convergence of the 1-norm, and because of the equivalence of vector norms, it is sufficient for the proof of the statement. We have

$$\left|\frac{1}{(k+1)^{2-2\gamma}} - \frac{1}{k^{2-2\gamma}}\right| \leqslant \frac{2(1-\gamma)}{k^{3-2\gamma}}$$

and

$$\sum_{i=1}^k\sum_{j=1}^k \mathbb{E}(Y_{i,q}Y_{j,q}) = \sum_{i=-q}^{k-1-q}\sum_{j=-q}^{k-1-q} \mathrm{Cov}(X_i, X_j) \leqslant \kappa k$$

for some constant κ according to Lemma 2.3.3.

Therefore,

$$\sum_{k=1}^{n-1}\left|\frac{1}{(k+1)^{2-2\gamma}}-\frac{1}{k^{2-2\gamma}}\right|\sum_{i=1}^{k}\sum_{j=1}^{k}\mathbb{E}(Y_{i,q}Y_{j,q})+2\sum_{k=1}^{n-1}k^{2\gamma-2}\,\mathbb{E}^{1/2}(Y_{k+1,q}^2)\left(\sum_{i,j=1}^{k}\mathbb{E}\left(Y_{i,q}Y_{j,q}\right)\right)^{1/2}$$
$$+2\sum_{k=0}^{n-1}k^{2\gamma-2}\,\mathbb{E}(Y_{k+1,q}^2)$$
$$\leqslant(2\kappa-2\kappa\gamma)\sum_{k=1}^{n-1}k^{2\gamma-2}+2(\kappa U_1)^{1/2}\sum_{k=1}^{n-1}k^{2\gamma-3/2}+2U_1\sum_{k=0}^{n-1}k^{2\gamma-2},$$

where U_1 is the upper boundary of $(\mathrm{Var}(X_n))_{n\in\mathbb{N}}$. The limit of the right hand side as $n\to\infty$ is finite, which completes the proof. We note the necessity of $\gamma<\frac{1}{4}$—otherwise the second term in the last expression would not be bounded. □

2.9.3 Lemma. *For a time-homogeneous INAR(p) process satisfying condition* $\mathbf{C}_0$ *and any* $\gamma<\frac{1}{4}$ *we have*

$$\max_{1\leqslant k\leqslant n}k^{\gamma-1}\left|\sum_{i=1}^{k}M_i\right|=\mathrm{O}_{\mathbb{P}}(1).$$

and

$$\max_{1\leqslant k\leqslant n}k^{\gamma-1}\left|\sum_{i=1}^{k}X_{i-1}M_i\right|=\mathrm{O}_{\mathbb{P}}(1).$$

Proof. We apply 2.9.1 in the same way as in the proof of Lemma 2.9.2 with $c_k=k^{\gamma-1}$ and $Y_i=M_i$. We note that the M_k are martingale differences, therefore any product $M_iM_j, i\neq j$ has zero mean. Furthermore, the sequence $(\mathrm{Var}\,M_k)_{k\in\mathbb{N}}$ is clearly bounded, and denoting its upper bound by U, we have

$$\sum_{k=1}^{n-1}\left|\frac{1}{(k+1)^{2-2\gamma}}-\frac{1}{k^{2-2\gamma}}\right|\sum_{i=1}^{k}\sum_{j=1}^{k}\mathbb{E}(Y_iY_j)+2\sum_{k=1}^{n-1}k^{2\gamma-2}\,\mathbb{E}^{1/2}(Y_{k+1}^2)\left(\sum_{i,j=1}^{k}\mathbb{E}\left(Y_iY_j\right)\right)^{1/2}$$
$$+2\sum_{k=0}^{n-1}k^{2\gamma-2}\,\mathbb{E}(Y_{k+1}^2)$$
$$\leqslant U(2-2\gamma)\sum_{k=1}^{n-1}k^{2\gamma-2}+2U\sum_{k=1}^{n-1}k^{2\gamma-3/2}+2U\sum_{k=0}^{n-1}k^{2\gamma-2},$$

whence the final steps of the proof are the same as in Lemma 2.9.2. □

2.9.4 Lemma. *Under the conditions of Theorem 2.7.1 we have*

$$\widehat{\boldsymbol{\theta}}_n-\widetilde{\boldsymbol{\theta}}=\mathrm{O}_{\mathbb{P}}(n^{-1/2}).$$

Proof. The difference can be decomposed in the following way:

(2.9.1)
$$n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}}) = (n^{-1}\boldsymbol{Q}_n)^{-1} n^{-1/2} \left[\sum_{k=1}^{\lfloor n\rho \rfloor} X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} - \boldsymbol{Q}_n \widetilde{\boldsymbol{Q}}^{-1} \left(\rho C' \begin{bmatrix} \boldsymbol{\alpha}' \\ \mu' \end{bmatrix} \right) \right.$$
$$\left. + \sum_{k=\lfloor n\rho \rfloor+1}^{n} X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} - \boldsymbol{Q}_n \widetilde{\boldsymbol{Q}}^{-1} \left((1-\rho) C'' \begin{bmatrix} \boldsymbol{\alpha}'' \\ \mu'' \end{bmatrix} \right) \right].$$

The first factor converges to $\widetilde{\boldsymbol{Q}}^{-1}$ stochastically, and will therefore be omitted from further calculations. The second factor has been split in two and only the first part will be analyzed in detail. The analysis of the second part is completely analogous. We split the first part in the second factor in the following way:

(2.9.2)
$$n^{-1/2} \left[\sum_{k=1}^{\lfloor n\rho \rfloor} X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} - \boldsymbol{Q}_n \widetilde{\boldsymbol{Q}}^{-1} \left(\rho \boldsymbol{Q}' \begin{bmatrix} \boldsymbol{\alpha}' \\ \mu' \end{bmatrix} \right) \right]$$
$$= n^{-1/2} \left(\sum_{k=1}^{\lfloor n\rho \rfloor} M_k' \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \right) + n^{-1/2} \left(\boldsymbol{Q}_{\lfloor n\rho \rfloor} - \rho \boldsymbol{Q}_n \widetilde{\boldsymbol{Q}}^{-1} \boldsymbol{Q}' \right) \begin{bmatrix} \boldsymbol{\alpha} \\ \mu' \end{bmatrix}.$$

The first term is
$$n^{-1/2} \left(\sum_{k=1}^{\lfloor n\rho \rfloor} \boldsymbol{Z}_k' \right),$$
which is asymptotically normal, and therefore $\mathrm{O}_{\mathbb{P}}(1)$ according to Theorem 2.5.3 (the same reasoning applies after the change, since Lindeberg's theorem is valid for triangular arrays as well). It remains to show that $n^{-1/2} \left(\boldsymbol{Q}_{\lfloor n\rho \rfloor} - \rho \boldsymbol{Q}_n \widetilde{\boldsymbol{Q}}^{-1} \boldsymbol{Q}' \right)$ is stochastically bounded. We decompose it in the following way:

(2.9.3)
$$n^{-1/2} \left(\boldsymbol{Q}_{\lfloor n\rho \rfloor} - \rho \boldsymbol{Q}_n \widetilde{\boldsymbol{Q}}^{-1} \boldsymbol{Q}' \right) = n^{-1/2} \left(\boldsymbol{Q}_{\lfloor n\rho \rfloor} - \mathbb{E}(\boldsymbol{Q}_{\lfloor n\rho \rfloor}) \right) + n^{-1/2} \left[\mathbb{E}(\boldsymbol{Q}_{\lfloor n\rho \rfloor}) - \lfloor n\rho \rfloor \boldsymbol{Q}' \right]$$
$$- n^{-1/2} \{ \rho [\boldsymbol{Q}_n - \mathbb{E}(\boldsymbol{Q}_n)] \widetilde{\boldsymbol{Q}}^{-1} \boldsymbol{Q}' \}$$
$$- n^{-1/2} \{ \rho [\mathbb{E}(\boldsymbol{Q}_n) - n\widetilde{\boldsymbol{Q}}] \widetilde{\boldsymbol{Q}}^{-1} \boldsymbol{Q}' \}$$
$$- n^{-1/2} \{ \rho n \boldsymbol{Q}' - \lfloor n\rho \rfloor \boldsymbol{Q}' \}.$$

The last term in (2.9.3) is deterministic and o(1). We know from (2.3.8) that the variances of the first and third terms are bounded. Denoting the common upper bound by K we have, from Markov's inequality, for all n,

$$\mathbb{P}\left(n^{-1}\left\|\boldsymbol{Q}_{\lfloor n\rho\rfloor}-\mathbb{E}(\boldsymbol{Q}_{\lfloor n\rho\rfloor})\right\|^2>a\right)<\frac{K}{a}\to 0 \text{ as } a\to\infty,$$

and similarly for the third term. Consequently, the first and third terms in (2.9.3) are $\mathrm{O}_{\mathbb{P}}(1)$. Recalling Lemma 2.3.1 we have

$$\begin{aligned}\left\|\mathbb{E}(\boldsymbol{Q}_{\lfloor n\rho\rfloor})-\lfloor n\rho\rfloor\boldsymbol{Q}'\right\| &= \left\|\sum_{k=1}^{\lfloor n\rho\rfloor}\left(\mathbb{E}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top-\boldsymbol{Q}'\right)\right\| \\ &\leqslant \sum_{k=1}^{\lfloor n\rho\rfloor}\left\|\mathbb{E}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_{k-1}\\ 1\end{bmatrix}^\top-\boldsymbol{Q}'\right\|\leqslant\sum_{k=1}^{\lfloor n\rho\rfloor}\pi^k=\mathrm{O}(1),\end{aligned}$$

because the matrices within the sum consist entirely of the entries of $\boldsymbol{X}_k^{\otimes 2}-\left(\widetilde{\boldsymbol{X}}'\right)^{\otimes 2}$. A similar calculation is valid for the fourth term. This implies the boundedness of the second and fourth terms of (2.9.3), hence our proof is complete. $\square$

2.9.5 Lemma. *Let* $a_n\to a>0$, $n\to\infty$ *and* $a_i>0$ *for all* $i\in\mathbb{N}$. *Then*

$$\min_{1\leqslant k\leqslant n}k^{-1}\sum_{i=n-k+1}^{n}a_i\to a,\quad n\to\infty.$$

Proof. First we note that for any $\varepsilon>0$ and sufficiently large n, we have

$$\min_{1\leqslant k\leqslant n}k^{-1}\sum_{i=n-k+1}^{n}a_i<a+\varepsilon.$$

This can be seen by choosing $k=1$ for every n. Now we show

$$\min_{1\leqslant k\leqslant n}k^{-1}\sum_{i=n-k+1}^{n}a_i>a-\varepsilon.$$

Let $\nu(\varepsilon)$ be the threshold index so that for $n>\nu(\varepsilon)$ we have $|a_n-a|<\frac{\varepsilon}{2}$. Let us denote by K the sum $\sum_{i=1}^{\nu(\varepsilon)}a_i$. Clearly,

$$\left|\min_{1\leqslant k\leqslant n-\nu(\varepsilon)}k^{-1}\sum_{i=n-k+1}^{n}a_i-a\right|<\frac{\varepsilon}{2}.$$

Furthermore, for any $n>k>n-\nu(\varepsilon)$ we have

$$k^{-1}\sum_{i=n-k+1}^{n}a_i>n^{-1}\sum_{i=\nu(\varepsilon)}^{n}a_i=\frac{n-\nu(\varepsilon)}{n}\left[(n-\nu(\varepsilon))^{-1}\sum_{i=\nu(\varepsilon)}^{n}a_i\right].$$

For sufficiently large n the first factor is close to 1, and the second factor is closer to a than $\frac{\varepsilon}{2}$ for every n. This suffices for the proof. $\square$

2.9.6 Lemma. *Let* $a_n \to a > 0,\ n \to \infty$ *and* $a_i > 0$ *for all* $i \in \mathbb{N}$. *Then*

$$\lim_{K\to\infty} \inf_{k\geqslant K} k^{-1} \sum_{i=1}^{k} a_i = a, \quad n \to \infty.$$

Proof. We only need to observe that convergence of a_n implies convergence in Cesaro mean as well, therefore, for a sufficiently large K and for all $k \geqslant K$ the average $k^{-1} \sum_{i=1}^{k} a_i$ is close to a. $\square$

2.9.7 Lemma. *For a time-homogeneous INAR(p) process satisfying condition* $\mathbf{C}_0$ *we have for any* $a > 0$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\max_{K\leqslant \ell\leqslant \lfloor n\rho\rfloor -1} \left\| \ell^{-1} \sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor} (\mathbb{E}(\boldsymbol{X}_{j-1}) - \boldsymbol{X}_{j-1}) \right\| > a \right) = 0$$

and

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\max_{K\leqslant \ell\leqslant \lfloor n\rho\rfloor -1} \left\| \ell^{-1} \sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor} \left(\mathbb{E}(\boldsymbol{X}_{j-1}^{\otimes 2}) - \boldsymbol{X}_{j-1}^{\otimes 2}\right) \right\| > a \right) = 0.$$

Similarly,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\max_{K\leqslant \ell\leqslant \lfloor n\rho\rfloor -1} \left\| \ell^{-1} \sum_{j=\lfloor n\rho\rfloor-\ell+1}^{\lfloor n\rho\rfloor} \begin{bmatrix} \boldsymbol{X}_{j-1} \\ 1 \end{bmatrix} M_j \right\| > a \right) = 0.$$

Proof. Similarly to the proof of Lemma 2.9.2 we will again employ Lemma 2.9.1 with $c_k = (K+k-1)^{-1}$ and $Y_{1,q} = \sum_{j=\lfloor n\rho\rfloor-K+1}^{\lfloor n\rho\rfloor} X_{j-q}$ and $Y_{i,q} = X_{\lfloor n\rho\rfloor-K+1-i-q}$ for $i \geqslant 2$ and $0 \leqslant q \leqslant p-1$.

By an easy calculation

$$\sum_{i,j=1}^{k} \mathbb{E}(Y_i Y_j) = \sum_{i,j=\lfloor n\rho\rfloor-K-k+1}^{\lfloor n\rho\rfloor} \mathbb{E}((\boldsymbol{X}_{i-1} - \mathbb{E}(\boldsymbol{X}_{i-1}))(\boldsymbol{X}_{j-1} - \mathbb{E}(\boldsymbol{X}_{j-1}))).$$

Therefore, applying the same estimations and notations as in the proof of Lemma 2.9.2 with $\gamma = 0$, we obtain the following upper limit for the probability in question:

$$2\kappa \sum_{\ell=K}^{\lfloor n\rho\rfloor-1} (\ell+1)^{-2} + U_1 \sum_{\ell=K}^{\lfloor n\rho\rfloor-1} (\ell+1)^{-3/2} + \frac{U_1}{K} + U_1 \sum_{\ell=K-1}^{\lfloor n\rho\rfloor-1} (\ell+1)^{-2}.$$

It is obvious that as $n \to \infty$ and then $K \to \infty$, the above expression converges to 0, which suffices for our proof. As in Lemma 2.9.2, for the second statement we merely take (ii) instead of (i) from Lemma 2.3.3.

For the third statement the arguments are the same, just as the proof of Lemma 2.9.3 is a simple analogue of Lemma 2.9.7. We note that $(M_n)_{n\in\mathbb{N}}$ is a martingale difference sequence, hence its elements are pairwise uncorrelated. Furthermore, $\mathrm{Var}(M_n)_{n\in\mathbb{N}}$ is bounded, which implies $\mathrm{Var}(M_1 + \ldots + M_n) = \mathrm{O}(n)$ immediately, and similarly for the other components of the vector to be summed. □

2.10 Illustration

Now we provide two real data examples of the use of our method. Since our model includes initial values, the series were not investigated in their full length, but the first p values were taken as the initial values $X_{-p+1}, \ldots, X_0$.

Our first example is the dataset of monthly polio cases in the US, as reported by the Centers for Disease Control and Prevention. It is available online at Hyndman (nd) and is 166 long. In Kang and Lee (2009) the authors found a significant decreasing trend in this series, while in Davis and Wu (2009) and Davis et al. (2000) the trend was found insignificant. It is widely agreed (see also Silva (2005)) that the underlying process is first-order, which is also supported by the partial autocorrelation function. Therefore we treated it as an INAR(1) process and calculated the CLS estimates given by (2.4.1). They were $\widehat{\alpha}_1 = 0.30646$ and $\widehat{\mu} = 0.94091$. The maximum of the absolute value of $\widehat{\mathcal{M}}^{(1)}_{166}$ was 1.2647 and the maximum of the absolute value of $\widehat{\mathcal{M}}^{(2)}_{166}$ was 1.1232. Applying the two-sided test simultaneously to the two parameters and requiring an overall significance level of 0.05, the critical value for each component is 1.48 (the individual significance levels are $1 - \sqrt{0.95} \approxeq 0.0253$), therefore, the null hypothesis is not rejected.

Our second example is a dataset of public drunkenness intakes in Minneapolis, also accessible at Hyndman (nd). This dataset is 139 long. After an examination of the partial autocorrelation function a seasonal INAR(12) model seems a rational choice, but with the assumption that only α_1 and α_{12} are nonzero (for another similar calculation, see the real data section in Barczy et al. (2011)). The estimates are

$$\begin{bmatrix} \widehat{\alpha}_1 \\ \widehat{\alpha}_{12} \\ \widehat{\mu} \end{bmatrix} = \left(\sum_{k=1}^{n} \begin{bmatrix} X_{k-1} \\ X_{k-12} \\ 1 \end{bmatrix} \begin{bmatrix} X_{k-1} \\ X_{k-12} \\ 1 \end{bmatrix}^{\top} \right)^{-1} \sum_{k=1}^{n} X_k \begin{bmatrix} X_{k-1} \\ X_{k-12} \\ 1 \end{bmatrix} = \begin{bmatrix} 0.8154 \\ 0.1419 \\ 9.6944 \end{bmatrix}.$$

The maxima of the absolute values of the respective components of $\widehat{\mathcal{M}}_n$ are 2.0333, 1.3497 and 1.5788. A comparison with the critical value of 1.545 (individual significance of approximately 0.017) results in the rejection of the null hypothesis. Based on $\left(\sum_{i=1}^{k} \widehat{M}_k^{(n)} X_{k-1}\right)_{k=1}^{139}$ our estimate for the change point is 41 (i.e., the 53rd entry in the original series). Repeating the procedure for the series before and after the change, the null hypothesis is accepted for both of them. For the series after the change, the CLS estimate of α_{12} is negative but an inspection of the partial autocorrelation function reveals that this series is more appropriately modeled as an INAR(1) process, for which the parameter estimates are $\widehat{\alpha}_1 = 0.8915$ and $\widehat{\mu} = 24.8429$ and the null hypothesis is accepted.

Chapter 3

General remarks about the Heston and Cox–Ingersoll–Ross models

In this brief chapter we will summarize some well-known properties of the Heston and Cox–Ingersoll–Ross models, which will be used repeatedly later on. Compared to the INAR(p) model, the results presented herein are deeper and require a more detailed knowledge of stochastic analysis; hence the decision to collect them here in one place, with reference to the papers and monographs where detailed proofs of these fundamental results can be found.

As a reminder, the Heston model is defined by

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma_1\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (\alpha - \beta Y_t)\,\mathrm{d}t + \sigma_2\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big), \end{cases} \qquad t \geqslant 0, \tag{3.0.1}$$

where $a \in \mathbb{R}_{++}$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1 \in \mathbb{R}_{++}$, $\sigma_2 \in \mathbb{R}_{++}$, $\varrho \in (-1, 1)$, and $(W_t, B_t)_{t\geqslant 0}$ is a 2-dimensional standard Wiener process. The Cox–Ingersoll–Ross process is the process Y in the definition, and in Chapter 4 we will use $\sigma := \sigma_1$.

3.1 Solutions and (conditional) means

The next proposition is about the existence and uniqueness of a strong solution of the SDE (3.0.1), see, e.g., Barczy and Pap (2016, Proposition 2.1).

3.1.1 Proposition. *Let (η_0, ζ_0) be a random vector independent of $(W_t, B_t)_{t\in\mathbb{R}_+}$ satisfying $\mathbb{P}(\eta_0 \in \mathbb{R}_+) = 1$. Then for all $a \in \mathbb{R}_{++}$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_{++}$, and $\varrho \in (-1, 1)$, there is a (pathwise) unique strong solution $(Y_t, X_t)_{t\in\mathbb{R}_+}$ of the SDE* (3.0.1) *such that $\mathbb{P}((Y_0, X_0) = (\eta_0, \zeta_0)) = 1$ and $\mathbb{P}(Y_t \in \mathbb{R}_+$ for all $t \in \mathbb{R}_+) = 1$. Further, for all $s, t \in \mathbb{R}_+$*

with $s \leqslant t$,

(3.1.1)
$$\begin{cases} Y_t = \mathrm{e}^{-b(t-s)}Y_s + a\int_s^t \mathrm{e}^{-b(t-u)}\,\mathrm{d}u + \sigma_1\int_s^t \mathrm{e}^{-b(t-u)}\sqrt{Y_u}\,\mathrm{d}W_u, \\ X_t = X_s + \int_s^t(\alpha - \beta Y_u)\,\mathrm{d}u + \sigma_2\int_s^t \sqrt{Y_u}\,\mathrm{d}(\varrho W_u + \sqrt{1-\varrho^2}B_u). \end{cases}$$

(3.1.2)
$$Y_t^2 = \mathrm{e}^{-2bt}Y_0^2 + \int_0^t \mathrm{e}^{-2b(t-u)}(2a+\sigma^2)Y_u\,\mathrm{d}u + 2\sigma\int_0^t \mathrm{e}^{-2b(t-u)}Y_u^{3/2}\,\mathrm{d}W_u, \quad t \in \mathbb{R}_+.$$

The conditional distribution of Y_t *on* Y_s, *where* $s < t$, *is noncentral chi-squared and we have*

$$\sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t^\eta) < \infty \qquad \textit{for all } \eta > 0. \tag{3.1.3}$$

Proof. By a theorem due to Yamada and Watanabe (see, e.g., Karatzas and Shreve, 1991, Proposition 5.2.13), the strong uniqueness holds for the first equation of (3.0.1). By Ikeda and Watanabe (1989, Example V.8.2, page 221), there is a (pathwise) unique non-negative strong solution $(Y_t)_{t\in\mathbb{R}_+}$ of the first equation of (3.0.1) with any initial value ξ independent of $(W_t)_{t\in\mathbb{R}_+}$ and satisfying $\mathbb{P}(\xi \in \mathbb{R}_+) = 1$. In this case we also have $\mathbb{P}(Y_t \in \mathbb{R}_+ \text{ for all } t \in \mathbb{R}_+) = 1$. From here it is a simple application of the Itô's formula for the process $(Y_t)_{t\in\mathbb{R}_+}$ that

$$\begin{aligned} \mathrm{d}(\mathrm{e}^{bt}Y_t) &= b\mathrm{e}^{bt}Y_t\,\mathrm{d}t + \mathrm{e}^{bt}\mathrm{d}Y_t = b\mathrm{e}^{bt}Y_t\,\mathrm{d}t + \mathrm{e}^{bt}\big((a-bY_t)\,\mathrm{d}t + \sigma\sqrt{Y_t}\,\mathrm{d}W_t\big) \\ &= a\mathrm{e}^{bt}\,\mathrm{d}t + \sigma\mathrm{e}^{bt}\sqrt{Y_t}\,\mathrm{d}W_t \end{aligned}$$

for all $t \in \mathbb{R}_+$. This implies the first equations in (3.1.1) – the rest can be obtained in the same manner. The noncentral chi-squared distribution is a well-known property of the process, and it can be found in the paper of Feller (1951). The property (3.1.3) is a direct consequence of this fact and the calculations can be found, e.g., in Ben Alaya and Kebaier (2013, Proposition 3). □

Next we present a result about the first moment of $(Y_t, X_t)_{t\in\mathbb{R}_+}$. For a proof, see, e.g., Barczy and Pap (2016, Proposition 2.2) together with (3.1.1) and Karatzas and Shreve (1991, Proposition 3.2.10).

3.1.2 Proposition. *Let* $(Y_t, X_t)_{t\in\mathbb{R}_+}$ *be the unique strong solution of the SDE* (3.0.1) *satisfying* $\mathbb{P}(Y_0 \in \mathbb{R}_+) = 1$ *and* $\mathbb{E}(Y_0) < \infty$, $\mathbb{E}(|X_0|) < \infty$. *Let us take* $s, t \in \mathbb{R}_+$ *such that* $s \leqslant t$.

In this case we have

$$\mathbb{E}(Y_t \,|\, \mathcal{F}_s) = \mathrm{e}^{-b(t-s)} Y_s + a \int_s^t \mathrm{e}^{-b(t-u)} \,\mathrm{d}u, \tag{3.1.4}$$

$$\begin{aligned} \mathbb{E}(X_t \,|\, \mathcal{F}_s) &= X_s + \int_s^t (\alpha - \beta\, \mathbb{E}(Y_u \,|\, \mathcal{F}_s)) \,\mathrm{d}u \\ &= X_s + \alpha(t-s) - \beta Y_s \int_s^t \mathrm{e}^{-b(u-s)} \,\mathrm{d}u - a\beta \int_s^t \left(\int_s^u \mathrm{e}^{-b(u-v)} \,\mathrm{d}v \right) \mathrm{d}u, \end{aligned} \tag{3.1.5}$$

and hence

$$\begin{bmatrix} \mathbb{E}(Y_t) \\ \mathbb{E}(X_t) \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{-bt} & 0 \\ -\beta \int_0^t \mathrm{e}^{-bu} \,\mathrm{d}u & 1 \end{bmatrix} \begin{bmatrix} \mathbb{E}(Y_0) \\ \mathbb{E}(X_0) \end{bmatrix} + \begin{bmatrix} \int_0^t \mathrm{e}^{-bu} \,\mathrm{d}u & 0 \\ -\beta \int_0^t \left(\int_0^u \mathrm{e}^{-bv} \,\mathrm{d}v \right) \mathrm{d}u & t \end{bmatrix} \begin{bmatrix} a \\ \alpha \end{bmatrix},$$

$$\mathbb{E}(Y_t^2) = \mathrm{e}^{-2bt}\, \mathbb{E}(Y_0^2) + \int_0^t (2a + \sigma^2) \left(\mathrm{e}^{-b(2t-u)}\, \mathbb{E}(Y_0) + a \int_0^u \mathrm{e}^{-b(2t-u-v)} \,\mathrm{d}v \right) \mathrm{d}u.$$

Consequently, if $b > 0$, *then*

$$\lim_{t\to\infty} \mathbb{E}(Y_t) = \frac{a}{b}, \qquad \lim_{t\to\infty} t^{-1}\, \mathbb{E}(X_t) = \alpha - \frac{\beta a}{b},$$

if $b = 0$, *then*

$$\lim_{t\to\infty} t^{-1}\, \mathbb{E}(Y_t) = a, \qquad \lim_{t\to\infty} t^{-2}\, \mathbb{E}(X_t) = -\frac{1}{2}\beta a,$$

if $b < 0$, *then*

$$\lim_{t\to\infty} \mathrm{e}^{bt}\, \mathbb{E}(Y_t) = \mathbb{E}(Y_0) - \frac{a}{b}, \qquad \lim_{t\to\infty} \mathrm{e}^{bt}\, \mathbb{E}(X_t) = \frac{\beta}{b}\, \mathbb{E}(Y_0) - \frac{\beta a}{b^2}.$$

Based on the asymptotic behavior of the expectations $(\mathbb{E}(Y_t), \mathbb{E}(X_t))$ as $t \to \infty$, we introduce a classification of the Heston model given by the SDE (3.0.1).

3.1.3 Definition. *Let* $(Y_t, X_t)_{t\in\mathbb{R}_+}$ *be the unique strong solution of the SDE* (3.0.1) *satisfying* $\mathbb{P}(Y_0 \in \mathbb{R}_+) = 1$. *We call* $(Y_t, X_t)_{t\in\mathbb{R}_+}$ *subcritical, critical or supercritical if* $b > 0$, $b = 0$ *or* $b < 0$, *respectively.*

3.2 Ergodic properties

The following result states the existence of a unique stationary distribution and the ergodicity for the CIR process $(Y_t)_{t\in\mathbb{R}_+}$ in the subcritical case. These statements are treated as evident in the literature, therefore we will omit the proof, and only note that the critical elements for it can be found, e.g., in Cox et al. (1985, Equation (20)), Li and Ma (2015, Theorem 2.6), Barczy et al. (2014a, Theorem 3.1 with $\alpha = 2$ and Theorem 4.1), or Jin et al. (2016, Corollaries 5.9 and 6.4). Only (3.2.4) can be considered as a slight improvement of the existing results.

3.2.1 Theorem. *Let* $a, b, \sigma_1 \in \mathbb{R}_{++}$. *Let* $(Y_t)_{t\in\mathbb{R}_+}$ *be the unique strong solution of the first equation of the SDE* (3.0.1) *satisfying* $\mathbb{P}(Y_0 \in \mathbb{R}_+) = 1$. *Then*

(i) $Y_t \xrightarrow{\mathcal{D}} Y_\infty$ *as* $t \to \infty$, *and the distribution of* Y_∞ *is given by*

$$\mathbb{E}(\mathrm{e}^{-\lambda Y_\infty}) = \left(1 + \frac{\sigma_1^2}{2b}\lambda\right)^{-2a/\sigma_1^2}, \qquad \lambda \in \mathbb{R}_+, \tag{3.2.1}$$

i.e., Y_∞ *has Gamma distribution with parameters* $2a/\sigma_1^2$ *and* $2b/\sigma_1^2$, *hence*

$$\mathbb{E}(Y_\infty) = \frac{a}{b}, \qquad \mathbb{E}(Y_\infty^2) = \frac{(2a+\sigma_1^2)a}{2b^2}, \qquad \mathbb{E}(Y_\infty^3) = \frac{(2a+\sigma_1^2)(a+\sigma_1^2)a}{2b^3}. \tag{3.2.2}$$

(ii) *supposing that the random initial value* Y_0 *has the same distribution as* Y_∞, *the process* $(Y_t)_{t\in\mathbb{R}_+}$ *is strictly stationary.*

(iii) *for all Borel measurable functions* $f : \mathbb{R} \to \mathbb{R}$ *such that* $\mathbb{E}(|f(Y_\infty)|) < \infty$, *we have*

$$\frac{1}{T}\int_0^T f(Y_s)\,\mathrm{d}s \xrightarrow{\text{a.s.}} \mathbb{E}(f(Y_\infty)) \qquad \text{as } T \to \infty, \tag{3.2.3}$$

$$\frac{1}{n}\sum_{i=0}^{n-1} f(Y_i) \xrightarrow{\text{a.s.}} \mathbb{E}(f(Y_\infty)) \qquad \text{as } n \to \infty. \tag{3.2.4}$$

Proof. Based on the references given before the theorem, we only need to show (3.2.4). By Corollary 2.7 in Jin et al. (2013), the tail σ-field $\bigcap_{t\in\mathbb{R}_+} \sigma(Y_s, s \geqslant t)$ of $(Y_t)_{t\in\mathbb{R}_+}$ is trivial for any initial distribution, i.e., the tail σ-field in question consists of events having probability 0 or 1 for any initial distribution. But since the tail σ-field of $(Y_t)_{t\in\mathbb{R}_+}$ is richer than that of $(Y_i)_{i\in\mathbb{Z}_+}$, the tail σ-field of $(Y_i)_{i\in\mathbb{Z}_+}$ is also trivial for any initial distribution.

Denoting the distribution of Y_0 and Y_∞ by ν and μ, respectively, let us introduce the distribution $\eta := (\mu + \nu)/2$. Let us introduce the following processes: $(Z_t)_{t\in\mathbb{R}_+}$, which is the CIR process with initial condition $Z_0 = \zeta_0$, where ζ_0 has the distribution μ; and $(U_t)_{t\in\mathbb{R}_+}$, which is the CIR process with initial condition $U_0 = \xi_0$, where ξ_0 has the distribution η.

We use Birkhoff's ergodic theorem (see, e.g., Theorem 8.4.1 in Dudley (2004)) in the usual setting: the probability space is $(\mathbb{R}^{\mathbb{Z}_+}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}_+}), \mathcal{L}((Z_i)_{i\in\mathbb{Z}_+}))$, where $\mathcal{L}((Z_i)_{i\in\mathbb{Z}_+})$ denotes the distribution of $(Z_i)_{i\in\mathbb{Z}_+}$, and the measure-preserving transformation T is the shift operator, i.e., $T((x_i)_{i\in\mathbb{Z}_+}) := (x_{i+1})_{i\in\mathbb{Z}_+}$ for $(x_i)_{i\in\mathbb{Z}_+} \in \mathbb{R}^{\mathbb{Z}_+}$ (the measure preserving property follows from (ii)). All invariant sets of T are included in the tail σ-field of the coordinate mappings π_i, $i \in \mathbb{Z}_+$, on $\mathbb{R}^{\mathbb{Z}_+}$, since for any invariant set A we have $A \in \sigma(\pi_0, \pi_1, \ldots)$, but as $T^k(A) = A$

for all $k \in \mathbb{N}$, it is also true that $A \in \sigma(\pi_k, \pi_{k+1}, \ldots)$ for all $k \in \mathbb{N}$. This implies that T is ergodic, since the tail σ-field is trivial. Hence we can apply the ergodic theorem for the function

$$g : \mathbb{R}^{\mathbb{Z}_+} \to \mathbb{R}, \qquad g((x_i)_{i\in\mathbb{Z}_+}) := f(x_0), \qquad (x_i)_{i\in\mathbb{Z}_+} \in \mathbb{R}^{\mathbb{Z}_+},$$

where f is given in (iii), to obtain

$$\frac{1}{n}\sum_{i=0}^{n-1} f(x_i) \to \int_{\mathbb{R}_+} f(x_0)\,\mu(\mathrm{d}x_0) \qquad \text{as } n \to \infty$$

for almost every $(x_i)_{i\in\mathbb{Z}_+} \in \mathbb{R}^{\mathbb{Z}_+}$ with respect to the measure $\mathcal{L}((Z_i)_{i\in\mathbb{Z}_+})$, and consequently

$$\frac{1}{n}\sum_{i=0}^{n-1} f(Z_i) \xrightarrow{\text{a.s.}} \mathbb{E}(f(Y_\infty)) \qquad \text{as } n \to \infty, \tag{3.2.5}$$

because the distribution of Y_∞ does not depend on the initial distribution. We introduce the following event, which is a tail event of $(Z_i)_{i\in\mathbb{Z}_+}$ and has probability 1 by (3.2.5):

$$C_Z := \left\{\omega \in \Omega : \frac{1}{n}\sum_{i=0}^{n-1} f(Z_i(\omega)) \to \mathbb{E}(f(Y_\infty)) \text{ as } n \to \infty\right\}.$$

The events C_Y and C_U are defined in a similar way and are tail events of $(Y_i)_{i\in\mathbb{Z}_+}$ and $(U_i)_{i\in\mathbb{Z}_+}$, respectively. Now we can write

$$\begin{aligned}
\mathbb{P}(C_U) &= \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\,\mathrm{d}\eta(x) \\
&= \frac{1}{2}\int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\,\mathrm{d}\mu(x) + \frac{1}{2}\int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\,\mathrm{d}\nu(x) \\
&\geqslant \frac{1}{2}\int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\,\mathrm{d}\mu(x) = \frac{1}{2}\int_0^\infty \mathbb{P}(C_Z \,|\, Z_0 = x)\,\mathrm{d}\mu(x) = \frac{1}{2}\,\mathbb{P}(C_Z) = \frac{1}{2}.
\end{aligned}$$

Here we used that $\mathbb{P}(C_U \,|\, U_0 = x) = \mathbb{P}(C_Z \,|\, Z_0 = x)$ μ-a.e. $x \in \mathbb{R}_+$, since the conditional probabilities on both sides depend only on the transition probability kernel of the CIR process given by the first SDE of (3.0.1) irrespective of the initial distribution. Further, we note that $\mathbb{P}(C_U \,|\, U_0 = x)$ is defined uniquely only η-a.e. $x \in \mathbb{R}_+$, but, by the definition of η, this means both μ-a.e. $x \in \mathbb{R}_+$, and ν-a.e. $x \in \mathbb{R}_+$, and similarly $\mathbb{P}(C_Z \,|\, Z_0 = x)$ is defined μ-a.e. $x \in \mathbb{R}_+$, so our equalities are valid. Thus, we have $\mathbb{P}(C_U) \geqslant \frac{1}{2}$. But since C_U is a tail event of $(U_i)_{i\in\mathbb{Z}_+}$, its probability must be either 0 or 1 (since the tail σ-field is trivial), hence $\mathbb{P}(C_U) = 1$. Hence

$$2 = \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\,\mathrm{d}\mu(x) + \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\,\mathrm{d}\nu(x) \leqslant \mu([0,\infty)) + \nu([0,\infty)) = 2,$$

yielding that

$$\int_0^\infty \mathbb{P}(C_U \mid U_0 = x)\,\mathrm{d}\mu(x) = \int_0^\infty \mathbb{P}(C_U \mid U_0 = x)\,\mathrm{d}\nu(x) = 1,$$

and the second equality is exactly (3.2.4) after we note that, by the same argument as above,

$$\int_0^\infty \mathbb{P}(C_U \mid U_0 = x)\,\mathrm{d}\nu(x) = \int_0^\infty \mathbb{P}(C_Y \mid Y_0 = x)\,\mathrm{d}\nu(x) = \mathbb{P}(C_Y).$$

With this our proof is complete. □

For a subcritical CIR process we can improve on the convergence stated in Proposition 3.1.1.

3.2.2 Lemma. *For a subcritical CIR process we have*

$$\lim_{t\to\infty} \mathbb{E}(Y_t) = \mathbb{E}(Y_\infty) = \frac{a}{b}, \qquad \lim_{t\to\infty} \mathbb{E}(Y_t^2) = \mathbb{E}(Y_\infty^2) = \frac{2a^2 + a^2\sigma^2}{2b^2}, \tag{3.2.6}$$

moreover,

$$\int_0^\infty |\,\mathbb{E}(Y_t) - \mathbb{E}(Y_\infty)|\,\mathrm{d}t < \infty, \qquad \int_0^\infty |\,\mathbb{E}(Y_t^2) - \mathbb{E}(Y_\infty^2)|\,\mathrm{d}t < \infty. \tag{3.2.7}$$

Proof. The first equalities are straightforward by taking expectations on both sides in Proposition 3.1.1 (we note that the stochastic integrals in question are indeed martingales due to (3.1.3)). From there, (3.2.6) is a question of elementary calculus: for the first equation we write

$$\lim_{t\to\infty}\left(\mathrm{e}^{-bt}\,\mathbb{E}(Y_0) + a\int_0^t \mathrm{e}^{-b(t-u)}\,\mathrm{d}u\right) = \lim_{t\to\infty} a\int_0^t \mathrm{e}^{-bv}\,\mathrm{d}v = a\int_0^\infty \mathrm{e}^{-bv}\,\mathrm{d}v = \frac{a}{b}. \tag{3.2.8}$$

For the second equation we observe

$$\begin{aligned}\int_0^t\int_0^u \mathrm{e}^{-b(2t-u-v)}\,\mathrm{d}v\mathrm{d}u &= \frac{1}{b}\left(\int_0^t (\mathrm{e}^{-2b(t-u)} - \mathrm{e}^{-b(2t-u)})\,\mathrm{d}u\right)\\ &= \frac{1}{b}\int_0^t \mathrm{e}^{-2bu}\,\mathrm{d}u + \frac{\mathrm{e}^{-bt}}{b}\int_0^t \mathrm{e}^{-bu}\,\mathrm{d}u\end{aligned} \tag{3.2.9}$$

and hence

$$\begin{aligned}&\lim_{t\to\infty}\left(\mathrm{e}^{-2bt}\,\mathbb{E}(Y_0^2) + \int_0^t (2a+\sigma^2)\left(\mathrm{e}^{-b(2t-u)}\,\mathbb{E}(Y_0) + a\int_0^u \mathrm{e}^{-b(2t-u-v)}\,\mathrm{d}v\right)\,\mathrm{d}u\right)\\ &= (2a+\sigma^2)\lim_{t\to\infty}\left(\mathbb{E}(Y_0)\mathrm{e}^{-bt}\int_0^t \mathrm{e}^{-bw}\,\mathrm{d}w + a\int_0^t\int_0^u \mathrm{e}^{-b(2t-u-v)}\,\mathrm{d}v\,\mathrm{d}u\right)\\ &= (2a+\sigma^2)\frac{1}{b}\int_0^\infty \mathrm{e}^{-2bw}\,\mathrm{d}w.\end{aligned} \tag{3.2.10}$$

For the first part of (3.2.7) we consider (keeping in mind (3.2.8))

$$|\,\mathbb{E}(Y_t) - \mathbb{E}(Y_\infty)| = \left| \mathrm{e}^{-bt}\,\mathbb{E}(Y_0) - a\int_t^\infty \mathrm{e}^{-bu}\,\mathrm{d}u \right| \leqslant \mathrm{e}^{-bt}\,\mathbb{E}(Y_0) + ab^{-1}\mathrm{e}^{-bt},$$

which yields the result immediately. For the second part, we combine (3.2.9) and (3.2.10) to obtain

$$\begin{aligned}|\,\mathbb{E}(Y_t^2) - \mathbb{E}(Y_\infty^2)| &= \Bigg| \mathrm{e}^{-2bt}\,\mathbb{E}(Y_0^2) + (2a+\sigma^2)\mathrm{e}^{-bt}\int_0^t \left(\mathbb{E}(Y_0)\mathrm{e}^{-bu} + \frac{1}{b}\mathrm{e}^{-bu}\right)\mathrm{d}u \\ &\qquad - \frac{1}{b}\int_t^\infty \mathrm{e}^{-2bu}\,\mathrm{d}u \Bigg| \\ &\leqslant \mathrm{e}^{-2bt}\,\mathbb{E}(Y_0^2) + (2a+\sigma^2)\mathrm{e}^{-bt}\left(\mathbb{E}(Y_0) + \frac{1}{b}\right)\frac{1}{b} + \frac{1}{2b^2}\mathrm{e}^{-2bt}.\end{aligned}$$

This yields the desired result immediately. □

3.3 Strong laws of large numbers and martingale CLT's

Finally, we state an appropriate version of the strong law of large numbers and a martingale central limit theorem both in continuous and discrete time, according to the continuous and discrete time observations in the following sections. Theorems 3.3.1 and 3.3.2 refer to the continuous case, while Theorems 3.3.3 and 3.3.4 refer to the discrete case.

3.3.1 Theorem. (Special case of Liptser and Shiryaev, 2001, Lemma 17.4) *Let the process $(W_t)_{t\in\mathbb{R}_+}$ be a standard Wiener process with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. Let $(\xi_t)_{t\in\mathbb{R}_+}$ be a measurable process adapted to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ such that*

$$\text{(3.3.1)}\qquad \mathbb{P}\left(\int_0^t \xi_u^2\,\mathrm{d}u < \infty\right) = 1, \quad t\in\mathbb{R}_+ \qquad \textit{and} \qquad \int_0^t \xi_u^2\,\mathrm{d}u \xrightarrow{\text{a.s.}} \infty \qquad \textit{as}\;\; t\to\infty.$$

Then

$$\text{(3.3.2)}\qquad \frac{\int_0^t \xi_u\,\mathrm{d}W_u}{\int_0^t \xi_u^2\,\mathrm{d}u} \xrightarrow{\text{a.s.}} 0 \qquad \textit{as}\;\; t\to\infty.$$

3.3.2 Theorem. (Special case of Jacod and Shiryaev, 2003, Corollary VIII.3.24.) *Let $(Y_t^n)_{t\in\mathbb{R}_+}$ be a series of locally square-integrable continuous martingales such that*

$$\langle Y^n\rangle_t \xrightarrow{\mathbb{P}} t, \quad t\in\mathbb{R}_+, \qquad \textit{as } n\to\infty.$$

Then $(Y^n)_{t\in\mathbb{R}_+} \xrightarrow{\mathcal{D}} (W_t)_{t\in\mathbb{R}_+}$, where $(W_t)_{t\in\mathbb{R}_+}$ is a standard Wiener process.

3.3.3 Theorem. (Shiryaev, 1989, Chapter VII, Section 5, Theorem 4) *Let us take a filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, \mathbb{P}\big)$. Let $(M_n)_{n\in\mathbb{N}}$ be a square-integrable martingale* *with respect to the filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *such that* $\mathbb{P}(M_0 = 0) = 1$ *and* $\mathbb{P}(\lim_{n\to\infty}\langle M\rangle_n = \infty) = 1$, *where* $(\langle M\rangle_n)_{n\in\mathbb{N}}$ *denotes the predictable quadratic variation process of* M. *Then*

$$\frac{M_n}{\langle M\rangle_n} \xrightarrow{\text{a.s.}} 0 \qquad \text{as} \quad n \to \infty.$$

3.3.4 Theorem. (Jacod and Shiryaev, 2003, Theorem VIII.3.33) *Let*

$$\{(\boldsymbol{M}_{n,k}, \mathcal{F}_{n,k}) : k = 0, 1, \ldots, k_n\}_{n\in\mathbb{N}}$$

be a sequence of d*-dimensional square-integrable martingales with* $\boldsymbol{M}_{n,0} = \boldsymbol{0}$ *such that there exists some symmetric, positive semi-definite non-random matrix* $\boldsymbol{D} \in \mathbb{R}^{d\times d}$ *such that*

$$\sum_{k=1}^{k_n} \mathbb{E}((\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})(\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})^\top \,|\, \mathcal{F}_{n,k-1}) \xrightarrow{\mathbb{P}} \boldsymbol{D} \qquad \text{as} \quad n \to \infty,$$

and for all $\varepsilon \in \mathbb{R}_{++}$,

$$\sum_{k=1}^{k_n} \mathbb{E}(\|\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1}\|^2 \mathbb{1}_{\{\|\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1}\| \geqslant \varepsilon\}} \,|\, \mathcal{F}_{n,k-1}) \xrightarrow{\mathbb{P}} 0 \qquad \text{as} \quad n \to \infty. \tag{3.3.3}$$

Then

$$\sum_{k=1}^{k_n} (\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1}) = \boldsymbol{M}_{n,k_n} \xrightarrow{\mathcal{D}} \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{D}) \qquad \text{as} \quad n \to \infty,$$

where $\mathcal{N}_d(\boldsymbol{0}, \boldsymbol{D})$ *denotes a* d*-dimensional normal distribution with mean vector* $\boldsymbol{0}$ *and covariance matrix* $\boldsymbol{D}$.

Chapter 4

Change detection in the continuous case

Just as for the INAR(p) process, we will only consider subcritical CIR processes, that is, when $b > 0$. We will use the first equation from (3.0.1), but replace σ_1 with σ, as that will be the single volatility parameter in this chapter – or, equivalently, we will base our calculations on (1.3.2). Based on the process definition, our statistical problem takes the following form: we would like to test the null hypothesis

$$\mathrm{H}_0 : (Y_t)_{t\in[0,T]} \text{ is the path of a CIR process}$$

against the alternative hypothesis

$$\begin{aligned}\mathrm{H}_\mathrm{A} : \exists \rho \in (0,1) : (Y_t)_{t\in[0,\rho T]} &\text{ is a CIR process with parameters } a = a',\ b = b',\text{ and}\\ (Y_t)_{t\in[\rho T,T]} &\text{ is a CIR process with parameters } a = a'',\ b = b'',\end{aligned}$$

where $a' > 0$, $a'' > 0$, $b' > 0$ and $b'' > 0$ with $(a', b') \neq (a'', b'')$.

4.1 Construction of parameter estimators

Our estimates will be motivated by the least-squares method, but we will not define them as solutions to a least-squares problem. Instead first we introduce least squares estimators based on low-frequency discrete time observations, then we will introduce our estimators as a formal analogy.

An LSE of (a, b) based on a discrete time observation $(Y_i)_{i\in\{0,1,\ldots,n\}}$, can be obtained by solving the extremum problem

$$\big(\widehat{a}_n^{\mathrm{D}}, \widehat{b}_n^{\mathrm{D}}\big) := \mathop{\arg\min}_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - Y_{i-1} - (a - bY_{i-1}))^2.$$

This is a simple exercise, which has the well-known solution

$$\begin{bmatrix} \widehat{a}_n^{\mathrm{D}} \\ \widehat{b}_n^{\mathrm{D}} \end{bmatrix} = \left(\sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^{\top} \right)^{-1} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ -Y_{i-1} \end{bmatrix} (Y_i - Y_{i-1}),$$

provided $n\sum_{i=1}^{n} Y_{i-1}^2 - \big(\sum_{i=1}^{n} Y_{i-1}\big)^2 > 0$. A heuristic motivation behind these estimators can be found, e.g., in Hu and Long (2007, p. 178). By a formal analogy, we introduce the estimator of (a, b) based on a continuous time observation $(Y_t)_{t\in[0,T]}$ as

$$\widehat{\boldsymbol{\theta}}_T := \begin{bmatrix} \widehat{a}_T \\ \widehat{b}_T \end{bmatrix} := \left(\int_0^T \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix}^{\top} \mathrm{d}s \right)^{-1} \int_0^T \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}Y_s, \tag{4.1.1}$$

provided that the inverse is defined, that is, $T\int_0^T Y_s^2\,\mathrm{d}s - \big(\int_0^T Y_s\,\mathrm{d}s\big)^2 > 0$, which is true a.s. – this is an easy exercise. To condense our notation, we will use

$$\boldsymbol{Q}_s := \int_0^s \begin{bmatrix} 1 \\ -Y_u \end{bmatrix} \begin{bmatrix} 1 \\ -Y_u \end{bmatrix}^{\top} \mathrm{d}u \qquad \text{and} \qquad \boldsymbol{d}_s := \int_0^s \begin{bmatrix} 1 \\ -Y_u \end{bmatrix} \mathrm{d}Y_u \tag{4.1.2}$$

4.1.1 Remark. The stochastic integral $\int_0^s Y_u\,\mathrm{d}Y_u$ is observable, since, by Itô's formula, we have $\mathrm{d}(Y_t^2) = 2Y_t\,\mathrm{d}Y_t + \sigma^2 Y_t\,\mathrm{d}t$, $t \in \mathbb{R}_+$, hence $\int_0^s Y_u\,\mathrm{d}Y_u = \frac{1}{2}\big(Y_s^2 - Y_0^2 - \sigma^2\int_0^s Y_u\,\mathrm{d}u\big)$.

4.1.2 Remark. These estimates are the same as $\widehat{a}^c(T)$ and $\widehat{b}^c(T)$ in Overbeck and Rydén (1997); this can be verified by a simple calculation. This means that even though they were introduced formally, our estimates have statistical meaning: they are the high-frequency limits of the conditional least squares estimates introduced for discrete observations; furthermore, they are strongly consistent. Overbeck (1998) also provides the more standard ML estimates, but we didn't choose them because they include the term $\int_0^T Y_s^{-1}\,\mathrm{d}s$, whose moments are rather difficult to handle and require additional constraints on the value of a.

Using the definition of the CIR process from (1.3.2) one can check that

$$\begin{bmatrix} \widehat{a}_T - a \\ \widehat{b}_T - b \end{bmatrix} = \boldsymbol{Q}_T^{-1} \begin{bmatrix} \sigma\int_0^T Y_s^{1/2}\,\mathrm{d}W_s \\ -\sigma\int_0^T Y_s^{3/2}\,\mathrm{d}W_s \end{bmatrix}. \tag{4.1.3}$$

In further calculations we will use

$$\widetilde{\boldsymbol{d}}_s := \sigma \begin{bmatrix} \int_0^s Y_u^{1/2} \, \mathrm{d}W_u \\ -\int_0^s Y_u^{3/2} \, \mathrm{d}W_u \end{bmatrix}. \tag{4.1.4}$$

4.2 Construction of the test process

Let us fix a time horizon $T \in \mathbb{R}_{++}$. Our test process will be introduced as a formal analogy to the efficient score vector, as is done in Gombay (2008). For this, we note that the estimator $(\widehat{a}_T, \widehat{b}_T)$ can also be represented as a solution to a least-squares problem, namely,

$$(\widehat{a}_T, \widehat{b}_T) = \operatorname*{arg\,min}_{(a,b)\in\mathbb{R}^2} \left(-\int_0^T (a - bY_s) \, \mathrm{d}Y_s + \frac{1}{2} \int_0^T (a - bY_s)^2 \, \mathrm{d}s \right).$$

This can be compared with the maximum likelihood estimator

$$(\widehat{a}_T^{MLE}, \widehat{b}_T^{MLE}) := \operatorname*{arg\,max}_{(a,b)\in\mathbb{R}^2} \Lambda_{a,b,T}(Y),$$

where the log-likelihood function $\Lambda_{a,b,T}$ has the form

$$\Lambda_{a,b,T}(Y) = \int_0^T \frac{a - bY_s}{\sigma^2 Y_s} \, \mathrm{d}Y_s - \frac{1}{2} \int_0^T \frac{(a - bY_s)^2}{\sigma^2 Y_s} \, \mathrm{d}s,$$

see, for example, Overbeck (1998). For the score vector, we take the partial derivatives of the log-likelihood function w.r.t. our parameters a and b, and we arrive, for time tT, $t \in [0, 1]$, at the process

$$\begin{bmatrix} \partial_a \Lambda_{a,b,tT}(Y) \\ \partial_b \Lambda_{a,b,tT}(Y) \end{bmatrix} = \begin{bmatrix} \int_0^{tT} \frac{\mathrm{d}Y_s}{\sigma^2 Y_s} - \int_0^{tT} \frac{(a-bY_s)\,\mathrm{d}s}{\sigma^2 Y_s} \\ -\int_0^{tT} \frac{\mathrm{d}Y_s}{\sigma^2} + \int_0^{tT} \frac{(a-bY_s)\,\mathrm{d}s}{\sigma^2} \end{bmatrix} = \int_0^{tT} \frac{1}{\sigma^2 Y_s} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}M_s,$$

where

$$M_s := Y_s - Y_0 - \int_0^s (a - bY_u) \, \mathrm{d}u = \sigma \int_0^s \sqrt{Y_u} \, \mathrm{d}W_u, \qquad s \in \mathbb{R}_+, \tag{4.2.1}$$

is a martingale. Instead of the maximum likelihood estimators we use $\widehat{\boldsymbol{\theta}}_T$ from (4.1.1), so, based on the similarity between the two least-squares problems, we will use the process

$$\int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}M_s = \widetilde{\boldsymbol{d}}_{tT} \tag{4.2.2}$$

as an analogue of the true efficient score vector process. The information contained in a continuous sample $(Y_u)_{u\in[0,tT]}$ is the quadratic variation of the efficient score vector process, namely,

$$\int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix}^\top \langle M \rangle_s \, \mathrm{d}s = \sigma^2 \int_0^{tT} \begin{bmatrix} Y_s & -Y_s^2 \\ -Y_s^2 & Y_s^3 \end{bmatrix} \mathrm{d}s =: \boldsymbol{I}_{tT}, \tag{4.2.3}$$

since $\langle M \rangle_s = \sigma^2 Y_s$, $s \in \mathbb{R}_+$. For each $s \in \mathbb{R}_+$, replacing the parameters by their estimates in M_s, we obtain an estimate $\widehat{M}_s^{(T)}$, i.e.,

$$\widehat{M}_s^{(T)} := Y_s - Y_0 - \int_0^s (\widehat{a}_T - \widehat{b}_T Y_u) \, \mathrm{d}u, \qquad s \in \mathbb{R}_+.$$

Our test process will be the estimated efficient score vector multiplied by the square root of the inverse of the information matrix, i.e.,

$$\widehat{\boldsymbol{\mathcal{M}}}_t^{(T)} := \boldsymbol{I}_T^{-1/2} \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}\widehat{M}_s^{(T)}, \qquad t \in [0, 1]. \tag{4.2.4}$$

This process can also be written in CUSUM form $\widehat{\boldsymbol{\mathcal{M}}}_t^{(T)} = \boldsymbol{I}_T^{-1/2} \boldsymbol{Q}_{tT} (\widehat{\boldsymbol{\theta}}_{tT} - \widehat{\boldsymbol{\theta}}_T)$, $t \in [0, 1]$. Indeed,

$$\begin{aligned} \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}\widehat{M}_s^{(T)} &= \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}Y_s - \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix}^\top \widehat{\boldsymbol{\theta}}_T \mathrm{d}s \\ &= \boldsymbol{Q}_{tT} \left(\boldsymbol{Q}_{tT}^{-1} \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}Y_s - \widehat{\boldsymbol{\theta}}_T \right). \end{aligned}$$

Under the null hypothesis the test process converges in distribution to a Brownian bridge, just like in Theorem 2.5.1.

4.2.1 Theorem. *Let $(Y_t)_{t \in \mathbb{R}_+}$ be a subcritical CIR process with $\mathbb{P}(Y_0 \in \mathbb{R}_+) = 1$. Then*

$$\left(\widehat{\boldsymbol{\mathcal{M}}}_t^{(T)} \right)_{t \in [0,1]} \xrightarrow{\mathcal{D}} (\boldsymbol{\mathcal{B}}_t)_{t \in [0,1]} \qquad \text{as } T \to \infty,$$

where $(\boldsymbol{\mathcal{B}}_t)_{t \in [0,1]}$ is a 2-dimensional standard Brownian bridge.

Proof. We have

$$\int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}\widehat{M}_s^{(T)} = \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}M_s - \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \left(\mathrm{d}M_s - \mathrm{d}\widehat{M}_s^{(T)} \right),$$

and

$$\int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \left(\mathrm{d}M_s - \mathrm{d}\widehat{M}_s^{(T)}\right) = \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \left(\widehat{a}_T - a - \big(\widehat{b}_T - b\big)Y_s\right)\mathrm{d}s$$

$$= \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix}^\top \begin{bmatrix} \widehat{a}_T - a \\ \widehat{b}_T - b \end{bmatrix} \mathrm{d}s = \boldsymbol{Q}_{tT}\boldsymbol{Q}_T^{-1}\widetilde{\boldsymbol{d}}_T,$$

with the notations from (4.1.2) and (4.1.4). Combining this with (4.2.2), for every $t \in [0,1]$,

$$\begin{aligned}\widehat{\boldsymbol{\mathcal{M}}}_t^{(T)} &= \boldsymbol{I}_T^{-1/2}\left(\widetilde{\boldsymbol{d}}_{tT} - \boldsymbol{Q}_{tT}\boldsymbol{Q}_T^{-1}\widetilde{\boldsymbol{d}}_T\right) = \boldsymbol{I}_T^{-1/2}\left(\widetilde{\boldsymbol{d}}_{tT} - t\widetilde{\boldsymbol{d}}_T\right) + \boldsymbol{I}_T^{-1/2}(t\boldsymbol{E}_2 - \boldsymbol{Q}_{tT}\boldsymbol{Q}_T^{-1})\widetilde{\boldsymbol{d}}_T \\ &= (T\boldsymbol{I})^{-1/2}\left(\widetilde{\boldsymbol{d}}_{tT} - t\widetilde{\boldsymbol{d}}_T\right) + \left((T^{-1}\boldsymbol{I}_T)^{-1/2} - \boldsymbol{I}^{-1/2}\right)T^{-1/2}\left(\widetilde{\boldsymbol{d}}_{tT} - t\widetilde{\boldsymbol{d}}_T\right) \\ &\quad + \boldsymbol{I}_T^{-1/2}(t\boldsymbol{E}_2 - \boldsymbol{Q}_{tT}\boldsymbol{Q}_T^{-1})\widetilde{\boldsymbol{d}}_T,\end{aligned}$$

where

$$\boldsymbol{I} := \sigma^2 \begin{bmatrix} \mathbb{E}(Y_\infty) & -\mathbb{E}(Y_\infty^2) \\ -\mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix}.$$

A simple consequence of the ergodic theorem is $T^{-1}\boldsymbol{I}_T \xrightarrow{\text{a.s.}} \boldsymbol{I}$. Consequently, Theorem 4.2.1 will follow from

$$\sup_{t\in[0,1]} (t\boldsymbol{E}_2 - \boldsymbol{Q}_{tT}\boldsymbol{Q}_T^{-1}) \xrightarrow{\mathbb{P}} 0 \qquad \text{as } T \to \infty, \tag{4.2.5}$$

and

$$\big(T^{-1/2}\,\widetilde{\boldsymbol{d}}_{tT}\big)_{t\in[0,1]} \xrightarrow{\mathcal{D}} (\boldsymbol{I}^{1/2}\,\boldsymbol{\mathcal{W}}_t)_{t\in[0,1]} \qquad \text{as } T \to \infty, \tag{4.2.6}$$

where $(\boldsymbol{\mathcal{W}}_t)_{t\in[0,1]}$ is a 2-dimensional standard Wiener process.

We begin by the proof of (4.2.6). The convergence is a simple consequence of the central limit theorem for continuous local martingales, see (Jacod and Shiryaev, 2003, Special case of Corollary VIII.3.24.). The process $\big(T^{-1/2}\,\widetilde{\boldsymbol{d}}_{tT}\big)_{t\in[0,1]}$ is a locally square-integrable martingale, therefore we only need to check the pointwise convergence of the quadratic variation. Using (iii) from Theorem 3.2.1 it is easy to show that, for every $t \in [0,1]$,

$$\frac{1}{T}\sigma^2 \int_0^{tT} \begin{bmatrix} Y_s & -Y_s^2 \\ -Y_s^2 & Y_s^3 \end{bmatrix} ds \xrightarrow{\text{a.s.}} \sigma^2 t \begin{bmatrix} \mathbb{E}(Y_\infty) & -\mathbb{E}(Y_\infty^2) \\ -\mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix} = t\boldsymbol{I}, \quad \text{as } T \to \infty.$$

For (4.2.5), introduce

$$\boldsymbol{Q} := \begin{bmatrix} 1 & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}$$

and note that due to Theorem 3.2.1 we have $T^{-1}\boldsymbol{Q}_T \xrightarrow{\text{a.s.}} \boldsymbol{Q}$. Now, first observe that

$$\|t\boldsymbol{E}_2 - \boldsymbol{Q}_{tT}\boldsymbol{Q}_T^{-1}\| \leqslant t\left\|\frac{\boldsymbol{Q}_T}{T} - \frac{\boldsymbol{Q}_{tT}}{tT}\right\| \left\|\left(\frac{\boldsymbol{Q}_T}{T}\right)^{-1}\right\|.$$

For this transformation to be sensible, we needed to extend $\frac{\boldsymbol{Q}_s}{s}$ continuously to $s = 0$, but this can be done since all components of $\frac{\boldsymbol{I}_s}{s}$ has a finite upper limit at 0 almost surely (i.e., the powers of Y_0). Since the last factor converges to $\|\boldsymbol{Q}^{-1}\|$ almost surely, for (4.2.5) it is sufficient to show that

$$\sup_{t\in[0,1]} t\left\|\frac{\boldsymbol{Q}_T}{T} - \frac{\boldsymbol{Q}_{tT}}{tT}\right\| \xrightarrow{\mathbb{P}} 0.$$

To exploit the almost sure convergence of $\frac{\boldsymbol{Q}_T}{T}$, we note that $\frac{\boldsymbol{Q}_T}{T} \xrightarrow{\text{a.s.}} \boldsymbol{Q}$ implies

$$\sup_{s\in[T,\infty)} \left\|\frac{\boldsymbol{Q}_s}{s} - \boldsymbol{Q}\right\| \xrightarrow{\text{a.s.}} 0$$

and thus also weakly. Now let us introduce $K := \sup_{s\in[T,\infty)} \left\|\frac{\boldsymbol{Q}_s}{s}\right\|$. This supremum is finite almost surely since $\frac{\boldsymbol{Q}_s}{s}$ is continuous on $\mathbb{R}_+$ and has a finite limit at infinity almost surely. Now we observe, for an arbitrary $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\in[0,1]} t\left\|\frac{\boldsymbol{Q}_T}{T} - \frac{\boldsymbol{Q}_{tT}}{tT}\right\| > \epsilon\right) &\leqslant \mathbb{P}\left(\sup_{0\leqslant t\leqslant \frac{\epsilon}{4K}\wedge 1} t\left\|\frac{\boldsymbol{Q}_T}{T} - \frac{\boldsymbol{Q}_{tT}}{tT}\right\| > \epsilon\right) \\
&\quad + \mathbb{P}\left(\sup_{\frac{\epsilon}{4K}\leqslant t\leqslant 1} t\left\|\frac{\boldsymbol{Q}_T}{T} - \frac{\boldsymbol{Q}_{tT}}{tT}\right\| > \epsilon\right) \\
&\leqslant \mathbb{P}\left(\frac{\epsilon}{4K}2K > \epsilon\right) \\
&\quad + \mathbb{P}\left(\sup_{\frac{\epsilon}{4K}\leqslant t\leqslant 1} \left(t\left\|\frac{\boldsymbol{Q}_T}{T} - \boldsymbol{Q}\right\| + \left\|\frac{\boldsymbol{Q}_{tT}}{tT} - \boldsymbol{Q}\right\|\right) > \epsilon\right) \\
&\leqslant 0 + \mathbb{P}\left(\left\|\frac{\boldsymbol{Q}_T}{T} - \boldsymbol{Q}\right\| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\sup_{\frac{\epsilon T}{4K}\leqslant s} \left\|\frac{\boldsymbol{Q}_s}{s} - \boldsymbol{Q}\right\| > \frac{\epsilon}{2}\right).
\end{aligned}$$

Dividing the last probability according to the value of K, we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\leqslant t\leqslant 1} t\left\|\frac{\boldsymbol{Q}_T}{T} - \frac{\boldsymbol{Q}_{tT}}{tT}\right\| > \epsilon\right) \\
&\quad\leqslant \mathbb{P}\left(\left\|\frac{\boldsymbol{Q}_T}{T} - \boldsymbol{Q}\right\| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\left\{\sup_{\frac{\epsilon T}{4K}\leqslant s} \left\|\frac{\boldsymbol{Q}_s}{s} - \boldsymbol{Q}\right\| > \frac{\epsilon}{2}\right\} \bigcap \left\{K \leqslant \sqrt{T}\right\}\right) + \mathbb{P}(K > \sqrt{T}),
\end{aligned}$$

and so,

$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant 1} t\left\|\frac{\boldsymbol{Q}_T}{T}-\frac{\boldsymbol{Q}_{tT}}{tT}\right\|>\epsilon\right)$$
$$\leqslant \mathbb{P}\left(\left\|\frac{\boldsymbol{Q}_T}{T}-\boldsymbol{Q}\right\|>\frac{\epsilon}{2}\right)+\mathbb{P}\left(\sup_{\frac{\epsilon\sqrt{T}}{4}\leqslant s}\left\|\frac{\boldsymbol{Q}_s}{s}-\boldsymbol{Q}\right\|>\frac{\epsilon}{2}\right)+\mathbb{P}\left(K>\sqrt{T}\right).$$

All three terms in the last expression tend to zero as $T\to\infty$, therefore (4.2.5) is proved. □

4.3 Testing procedures

Let us denote the components of the test process $(\widehat{\boldsymbol{\mathcal{M}}}_t^{(T)})_{t\in[0,1]}$ by $(\widehat{\boldsymbol{\mathcal{M}}}_{t,i}^{(T)})_{t\in[0,1]}$, $i\in\{1,2\}$. Based on Theorem 4.2.1, we can develop the following tests with a significance level of α:

Test 1 (one-sided): if it is clear that, in case of a change, $a'<a''$, reject H_0 if the infimum of $(\widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)})_{t\in[0,1]}$ is smaller than $C_1(\alpha)$, where $C_1(\alpha)$ can be obtained from the distribution of the infimum of a standard Brownian bridge. The same test can be applied to the supremum (for $a'>a''$) and to $(\widehat{\boldsymbol{\mathcal{M}}}_{t,2}^{(T)})_{t\in[0,1]}$ (for a change in b).

Test 2 (two-sided): reject H_0 if the supremum of $(|\widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)}|)_{t\in[0,1]}$ is greater than $C_2(\alpha)$, where $C_2(\alpha)$ can be obtained from the distribution of the supremum of the absolute value of standard Brownian bridge. The same test can be applied to $(|\widehat{\boldsymbol{\mathcal{M}}}_{t,2}^{(T)}|)_{t\in[0,1]}$ for a change in b.

Naturally, the test for a and b can be applied simultaneously, in which case the significance levels for the individual tests have to be modified accordingly, in order to produce an overall significance level of α – for more details, see subsection 2.5.1.

4.4 Asymptotic behavior under the alternative hypothesis

Before stating our results under the alternative hypothesis, we need more closely to examine the ergodicity results that we can use. Let us take two parameter vectors: $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$. Furthermore, we take two random variables, Y_∞' and Y_∞'', such that they are distributed according to the stationary distributions corresponding to $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$, respectively. Let us take a process $(Y_t)_{t\in\mathbb{R}_+}$ such that it evolves according to (1.3.2) with parameters $\boldsymbol{\theta}'$ until $t=\rho T$ and with parameters $\boldsymbol{\theta}''$ thereafter. This implies that the calculation of the martingale M should also be different according to whether $t<\rho T$. We will thus use

$$M_s':=Y_s-Y_0-\int_0^s(a'-b'Y_u)\,\mathrm{d}u \qquad \text{and} \qquad M_s'':=Y_s-Y_{\rho T}-\int_{\rho T}^s(a''-b''Y_u)\,\mathrm{d}u$$

for $s<\rho T$ and $s\geqslant\rho T$, respectively. We would like to apply the ergodic theorem (i.e., Theorem 3.2.1) separately to the process before and after the change-point (i.e., ρT). However, we cannot do this directly for the second part because the initial distribution may depend on T. However, we do have the following parallel of (2.6.1).

4.4.1 Lemma. *For a CIR process with the above conditions*

$$\frac{1}{T-\rho T}\int_{\rho T}^{T} g(Y_t)\,\mathrm{d}t \xrightarrow{\mathbb{P}} \mathbb{E}(g(\widetilde{Y}'')), \tag{4.4.1}$$

where $g : \mathbb{R}_+ \to \mathbb{R}$ *with* $\mathbb{E}(|g(\widetilde{Y})|) < \infty$.

Proof. For an arbitrary $\varepsilon > 0$

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{T-\rho T}\int_{\rho T}^{T} g(Y_t)\,\mathrm{d}t - \mathbb{E}(g(\widetilde{Y}''))\right| > \varepsilon\right)\\
&= \int_{\mathbb{R}_+} \mathbb{P}\left(\left|\frac{1}{T-\rho T}\int_{\rho T}^{T} g(Y_t)\,\mathrm{d}t - \mathbb{E}(g(\widetilde{Y}''))\right| > \varepsilon \,\middle|\, Y_{\rho T} = x\right) \mathrm{d}P_Y^{\rho T}(x)\\
&\leqslant \left\|P_Y^{\rho T} - P^*\right\| + \int_{\mathbb{R}_+} \mathbb{P}\left(\left|\frac{1}{T-\rho T}\int_{\rho T}^{T} g(Y_t)\,\mathrm{d}t - \mathbb{E}(g(\widetilde{Y}''))\right| > \varepsilon \,\middle|\, Y_{\rho T} = x\right) \mathrm{d}P^*(x),
\end{aligned}$$

where P^* is the distribution of $\widetilde{Y}'$, $P_Y^{\rho T}$ is the distribution of $Y_{\rho T}$ and $\|\cdot\|$ is the total variation norm. The first term converges to zero because the CIR process is positive Harris recurrent (Jin et al., 2013, Theorem 2.5). This implies ergodicity by Meyn and Tweedie (1993, Theorem 6.1), since in this case the 1-skeleton (i.e., the process $(Y_i)_{i\in\mathbb{Z}_+}$) is clearly irreducible because the support of the distribution of Y_1 conditionally on Y_0 is $\mathbb{R}_+$. In the second term the measure is finite, while the integrand is bounded by 1 and converges to zero pointwise, therefore (4.4.1) is proved by the Lebesgue Dominated Convergence Theorem. □

The same line of reasoning can be used to apply Theorems 3.3.1 and 3.3.2 after the point of change. Let us now introduce

$$\boldsymbol{d}_{[a,b]} := \begin{bmatrix} \int_a^b 1\mathrm{d}Y_s \\ -\int_a^b Y_s\mathrm{d}Y_s \end{bmatrix}, \qquad \boldsymbol{Q}_{[a,b]} := \begin{bmatrix} \int_a^b 1\,\mathrm{d}s & -\int_a^b Y_s\,\mathrm{d}s \\ -\int_a^b Y_s\,\mathrm{d}s & \int_a^b Y_s^2\,\mathrm{d}s \end{bmatrix}.$$

With these notations,

$$\widehat{\boldsymbol{\theta}}_T = \left(\boldsymbol{Q}_{[0,\rho T]} + \boldsymbol{Q}_{[\rho T,T]}\right)^{-1}(\boldsymbol{d}_{[0,\rho T]} + \boldsymbol{d}_{[\rho T,T]}).$$

With the help of the ergodic theorem, we can see that this quantity has a finite weak limit:

$$\widetilde{\boldsymbol{\theta}} := \begin{bmatrix} \widetilde{a} \\ \widetilde{b} \end{bmatrix} := (\rho\boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}'')^{-1}\left(\rho\boldsymbol{Q}'\boldsymbol{\theta}' + (1-\rho)\boldsymbol{Q}''\boldsymbol{\theta}''\right),$$

where

$$\boldsymbol{Q}' := \begin{bmatrix} 1 & -\mathbb{E}(Y'_\infty) \\ -\mathbb{E}(Y'_\infty) & \mathbb{E}((Y'_\infty)^2) \end{bmatrix}, \qquad \boldsymbol{Q}'' := \begin{bmatrix} 1 & -\mathbb{E}(Y''_\infty) \\ -\mathbb{E}(Y''_\infty) & \mathbb{E}((Y''_\infty)^2) \end{bmatrix}.$$

Furthermore, the information matrix will have a weak limit in this case, namely,

(4.4.2)

$$\frac{\boldsymbol{I}_T}{T} \xrightarrow{\mathbb{P}} \widetilde{\boldsymbol{I}} := \sigma^2 \left(\rho \begin{bmatrix} \mathbb{E}(Y'_\infty) & -\mathbb{E}((Y'_\infty)^2) \\ -\mathbb{E}((Y'_\infty)^2) & \mathbb{E}((Y'_\infty)^3) \end{bmatrix} + (1-\rho) \begin{bmatrix} \mathbb{E}(Y''_\infty) & -\mathbb{E}((Y''_\infty)^2) \\ -\mathbb{E}((Y''_\infty)^2) & \mathbb{E}((Y''_\infty)^3) \end{bmatrix} \right)$$

4.5 Asymptotic consistence of the test

Armed with these tools, we can derive the asymptotic behavior of the supremum of the components $(\widehat{\boldsymbol{\mathcal{M}}}^{(T)}_{t,i})_{t\in[0,1]}$, $i = 1, 2$, of the test process.

4.5.1 Theorem. *Let us suppose that $\boldsymbol{\theta}$ changes from $\boldsymbol{\theta}'$ to $\boldsymbol{\theta}''$ at time ρT, where $\rho \in (0, 1)$, and both $\boldsymbol{\theta}' > 0$ and $\boldsymbol{\theta}'' > 0$ componentwise. Let us take $i \in \{1, 2\}$, and then define*

$$\psi_i := \mathbf{1}_i^\top \widetilde{\boldsymbol{I}}^{-1/2} ((\rho \boldsymbol{Q}')^{-1} + ((1-\rho)\boldsymbol{Q}'')^{-1})^{-1} (\boldsymbol{\theta}' - \boldsymbol{\theta}'').$$

If $\psi_i > 0$, then we have

$$\sup_{t\in[0,T]} \widehat{\boldsymbol{\mathcal{M}}}^{(T)}_{t,i} = T^{1/2}\psi_i + \mathrm{o}_{\mathbb{P}}\left(T^{1/2}\right).$$

On the other hand, if $\psi_i < 0$, we have

$$\inf_{t\in[0,T]} \widehat{\boldsymbol{\mathcal{M}}}^{(T)}_{t,i} = T^{1/2}\psi_i + \mathrm{o}_{\mathbb{P}}\left(T^{1/2}\right).$$

4.5.2 Remark. We can easily see that in the special case of Theorem 4.5.1 when only a changes from $a' > 0$ to $a'' > 0$, we have

$$\psi_1 = (a' - a'')\mathbf{1}_1^\top \widetilde{\boldsymbol{I}}^{-1/2} ((\rho \boldsymbol{Q}')^{-1} + ((1-\rho)\boldsymbol{Q}'')^{-1})^{-1} \mathbf{1}_1,$$

and similarly, if only b changes from b' to b'', we have

$$\psi_2 = (b' - b'')\mathbf{1}_2^\top \widetilde{\boldsymbol{I}}^{-1/2} ((\rho \boldsymbol{Q}')^{-1} + ((1-\rho)\boldsymbol{Q}'')^{-1})^{-1} \mathbf{1}_2.$$

4.5.3 Corollary. *If $\psi_i \neq 0$, then for $\widehat{\boldsymbol{\mathcal{M}}}^{(T)}_{t,i}$ the two-sided test and the appropriate one-sided test described in 4.3 are asymptotically weakly consistent, that is, $\mathbb{P}(\mathrm{H}_0$ is rejected$) \to 1$ as $T \to \infty$.*

4.5.4 Remark. This theorem does not prove the consistence of our test if $\psi_i = 0$. This degenerate case is indeed possible, but characterizing it is not easy since the matrices $\boldsymbol{Q}'$ and $\boldsymbol{Q}''$ depend on $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$ in a nontrivial manner. What we can easily see is that when the change occurs in the same direction in a and b (including the case when only one of them changes), this is not a problem. If the changes are in opposite directions, however, then there is a point where they "cancel out" and we can't prove Corollary 4.5.3. It can also be easily checked that

$$\widetilde{\boldsymbol{I}}^{-1/2}((\rho\boldsymbol{Q}')^{-1} + ((1-\rho)\boldsymbol{Q}'')^{-1})^{-1}$$

is positive definite, so if only one parameter changes, the sign of the corresponding ψ_i (i.e., ψ_1 for a an ψ_2 for b) depends on the direction: it is negative in case of an upwards change and positive in case of a downwards change. This gives us the possibility to design one-sided tests, as described in section 4.3.

Proof. We will only prove for $i = 1$ and $\psi_1 > 0$, the other cases are completely analogous. By the definition of $\widehat{\boldsymbol{\mathcal{M}}}_t^{(T)}$, we have

$$\begin{aligned} T^{-1/2}\widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)} &= T^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_0^{tT}\begin{bmatrix}1\\-Y_s\end{bmatrix}\mathrm{d}\widehat{M}_s^{(T)} \\ &\quad + T^{-1}\mathbf{1}_1\left(\left(\frac{\boldsymbol{I}_T}{T}\right)^{-1/2} - \widetilde{\boldsymbol{I}}^{-1/2}\right)\int_0^{tT}\begin{bmatrix}1\\-Y_s\end{bmatrix}\mathrm{d}\widehat{M}_s^{(T)}. \end{aligned} \tag{4.5.1}$$

We need to show that the supremum of this expression is $\psi_1 + \mathrm{o}_{\mathbb{P}}(1)$. It is easily verifiable that

$$\begin{aligned} S(0,tT) :&= T^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_0^{tT}\begin{bmatrix}1\\-Y_s\end{bmatrix}\mathrm{d}\widehat{M}_s^{(T)} \\ &= T^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_0^{tT}\begin{bmatrix}1\\-Y_u\end{bmatrix}\mathrm{d}M_u' \\ &\quad + T^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_0^{tT}\begin{bmatrix}1 & -\mathbb{E}(Y_u)\\ -\mathbb{E}(Y_u) & \mathbb{E}(Y_u^2)\end{bmatrix}(\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}})\,\mathrm{d}u \\ &\quad + T^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_0^{tT}\begin{bmatrix}0 & \mathbb{E}(Y_u) - Y_u\\ \mathbb{E}(Y_u) - Y_u & Y_u^2 - \mathbb{E}(Y_u^2)\end{bmatrix}(\boldsymbol{\theta}' - \widehat{\boldsymbol{\theta}}_T)\,\mathrm{d}u \\ &\quad + T^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_0^{tT}\begin{bmatrix}1 & -\mathbb{E}(Y_u)\\ -\mathbb{E}(Y_u) & \mathbb{E}(Y_u^2)\end{bmatrix}(\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_T)\,\mathrm{d}u \\ &=: S_1(0,tT,\boldsymbol{\theta}') + S_2(0,tT,\boldsymbol{\theta}') + S_3(0,tT,\boldsymbol{\theta}') + S_4(0,tT), \end{aligned} \tag{4.5.2}$$

and we can also see
(4.5.3)

$$T^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_0^{tT}\begin{bmatrix}1\\-Y_s\end{bmatrix}\mathrm{d}\widehat{M}_s^{(T)} = S_1(0,tT,\boldsymbol{\theta}'') + S_2(0,tT,\boldsymbol{\theta}'') + S_3(0,tT,\boldsymbol{\theta}'') + S_4(0,tT)$$

as well. Let us now introduce $S_i(a,b,\theta) := S_i(0,b,\theta) - S_i(0,a,\theta)$ for $i = 1,2,3$ (this corresponds to taking the lower limit of the integral in the definition of $S_i(0,b,\theta)$ as a instead of zero). Now we can write

$$\begin{aligned}\left|\sup_{t\in[0,1]} S(0,tT) - \psi_1\right| \leqslant & \sup_{t\in[0,1]}\left|S_1(0,(\rho\wedge t)T,\boldsymbol{\theta}') + S_1((\rho\wedge t)T,tT,\boldsymbol{\theta}'')\right| \\ & + \left|\sup_{t\in[0,1]}\big(S_2(0,(\rho\wedge t)T,\boldsymbol{\theta}') + S_2((\rho\wedge t)T,tT,\boldsymbol{\theta}'')\big) - \psi_1\right| \\ & + \sup_{t\in[0,1]}\left|S_3(0,(\rho\wedge t)T,\boldsymbol{\theta}') + S_3((\rho\wedge t)T,tT\boldsymbol{\theta}'')\right| + \sup_{t\in[0,1]}\left|S_4(0,tT)\right|.\end{aligned}$$

The first term is $\mathrm{o}_{\mathbb{P}}(1)$ according to Lemma 4.7.4 with $\gamma = 0$, the third term by Lemma 4.7.3 and the fourth term by Lemma 4.7.5.

Now we turn to the second term. Let us notice that $\boldsymbol{Q}'$ and $\boldsymbol{Q}''$ are both symmetric, which we will exploit repeatedly. Clearly,

$$\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}} = (1-\rho)(\rho\boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}'')^{-1}\boldsymbol{Q}''(\boldsymbol{\theta}' - \boldsymbol{\theta}''),$$

and so

$$\begin{aligned}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\begin{bmatrix}1 & -\mathbb{E}(Y'_\infty)\\ -\mathbb{E}(Y'_\infty) & \mathbb{E}((Y'_\infty)^2)\end{bmatrix}(\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}}) &= \mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\boldsymbol{Q}'(1-\rho)(\rho\boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}'')^{-1}\boldsymbol{Q}''(\boldsymbol{\theta}' - \boldsymbol{\theta}'') \\ &= \frac{\psi_1}{\rho},\end{aligned}$$

and similarly,

$$\begin{aligned}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\begin{bmatrix}1 & -\mathbb{E}(Y''_\infty)\\ -\mathbb{E}(Y''_\infty) & \mathbb{E}((Y''_\infty)^2)\end{bmatrix}(\boldsymbol{\theta}'' - \widetilde{\boldsymbol{\theta}}) &= \mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\boldsymbol{Q}''(1-\rho)(\rho\boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}'')^{-1}\boldsymbol{Q}''(\boldsymbol{\theta}' - \boldsymbol{\theta}'') \\ &= \frac{\psi_1}{1-\rho}.\end{aligned}$$

We have exploited the fact that by an easy calculation,

$$\left((\rho\boldsymbol{Q}')^{-1} + ((1-\rho)\boldsymbol{Q}'')^{-1}\right)^{-1} = \rho(1-\rho)\left(\boldsymbol{Q}'\left(\rho\boldsymbol{Q}' + (1-\rho)\boldsymbol{Q}''\right)^{-1}\boldsymbol{Q}''\right).$$

Now we can write

$$\begin{aligned}\left|\sup_{t\in[0,1]}\bigl(S_2(0,(\rho\wedge t)T,\boldsymbol{\theta}')+S_2((\rho\wedge t)T,tT,\boldsymbol{\theta}'')-\psi_1\bigr)\right| &\leqslant \left|\sup_{t\in[0,1]}\left(\frac{\rho\wedge t}{\rho}-\frac{(t-\rho)^+}{(1-\rho)}-1\right)\psi_1\right| \\ &+ T^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sup_{t\in[0,1]}\left|\int_0^{(\rho T)\wedge t}\begin{bmatrix}0 & \mathbb{E}(Y'_\infty)-\mathbb{E}(Y_u)\\ \mathbb{E}(Y'_\infty)-\mathbb{E}(Y_u) & \mathbb{E}(Y_u^2)-\mathbb{E}(Y'^2_\infty)\end{bmatrix}^\top(\boldsymbol{\theta}'-\widetilde{\boldsymbol{\theta}})\,\mathrm{d}u\right| \\ &+ T^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\sup_{t\in[0,1]}\left|\int_0^{(\rho T)\wedge t}\begin{bmatrix}0 & \mathbb{E}(Y'_\infty)-\mathbb{E}(Y_u)\\ \mathbb{E}(Y'_\infty)-\mathbb{E}(Y_u) & \mathbb{E}(Y_u^2)-\mathbb{E}(Y'^2_\infty)\end{bmatrix}^\top(\boldsymbol{\theta}'-\widetilde{\boldsymbol{\theta}})\,\mathrm{d}u\right|.\end{aligned}$$

The first term is obviously zero, with the supremum attained at $t=\rho$; the other two terms converge to zero by (3.2.7). This concludes the proof for the first term in (4.5.1).

All that remains is showing that the second term in (4.5.1) is $\mathrm{o}_{\mathbb{P}}(1)$. To see this, consider, with the L_1-norm $\|\cdot\|$, and its induced matrix norm $\|\cdot\|_*$,

$$\begin{aligned}&\sup_{t\in[0,1]}T^{-1}\mathbf{1}_1\left(\left(\frac{\boldsymbol{I}_T}{T}\right)^{-1/2}-\widetilde{\boldsymbol{I}}^{-1/2}\right)\int_0^{tT}\begin{bmatrix}1\\-Y_s\end{bmatrix}\mathrm{d}\widehat{M}_s^{(T)} \\ &\qquad\leqslant\left\|\left(\frac{\boldsymbol{I}_T}{T}\right)^{-1/2}-\widetilde{\boldsymbol{I}}^{-1/2}\right\|_*\left\|\widetilde{\boldsymbol{I}}^{1/2}\right\|_*\sup_{t\in[0,1]}\left\|T^{-1}\widetilde{\boldsymbol{I}}^{-1/2}\int_0^{tT}\begin{bmatrix}1\\-Y_s\end{bmatrix}\mathrm{d}\widehat{M}_s^{(T)}\right\|,\end{aligned}$$

which is clearly $\mathrm{o}_{\mathbb{P}}(1)$ since the first factor is $\mathrm{o}_{\mathbb{P}}(1)$ (note that $\widetilde{\boldsymbol{I}}$ is invertible, hence we can use (4.4.2) and the continuous mapping theorem), the second factor is finite, and the third has just been shown to be $K+\mathrm{o}_{\mathbb{P}}(1)$ for some constant K. This completes the proof. □

4.5.5 Remark. It is apparent that the structure of the proof is essentially the same as for Theorem 2.7.1 and some arguments are even simpler – e.g., the change occurs exactly at ρT so we do not need to deal with the fractional part of the change point separately, as in the analysis of the second term of (2.7.4). These simplifications must be weighed against the need to include stochastic analysis and continuous martingale theory in our investigations.

4.6 Estimation of the change point

The natural estimate of the change point when a downward change in a is being tested, so when $\psi_1>0$, is $\widehat{\rho}_T T$, where

$$\widehat{\rho}_T := \inf\Bigl\{t\in[0,1]:\widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)}=\sup_{s\in[0,1]}\widehat{\boldsymbol{\mathcal{M}}}_{s,1}^{(T)}\Bigr\}. \tag{4.6.1}$$

Clearly, this is a well-defined, finite quantity, since $\widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)}$ has continuous trajectories almost surely. If we are looking for an upward change in a, i.e., $a'<a''$, when $\psi_1<0$, then the appropriate estimate is

$$\inf\Big\{t \in [0,1] : \widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)} = \inf_{s\in[0,1]} \widehat{\boldsymbol{\mathcal{M}}}_{s,1}^{(T)}\Big\}.$$

For a change in b, the appropriate estimates are

$$\inf\Big\{t \in [0,1] : \widehat{\boldsymbol{\mathcal{M}}}_{t,2}^{(T)} = \sup_{s\in[0,1]} \widehat{\boldsymbol{\mathcal{M}}}_{s,2}^{(T)}\Big\} \quad \text{and} \quad \inf\Big\{t \in [0,1] : \widehat{\boldsymbol{\mathcal{M}}}_{t,2}^{(T)} = \inf_{s\in[0,1]} \widehat{\boldsymbol{\mathcal{M}}}_{s,2}^{(T)}\Big\},$$

for a downward and upward change, respectively, corresponding to the different tests described in 4.3. We can define the estimate based on the two-sided test as well, but that will eventually reduce to one of these four cases, according to the sign of the appropriate ψ.

4.6.1 Theorem. *Under the assumptions of Theorem 4.5.1, if $\psi_i \neq 0$, then for the appropriate change-point estimate we have $\widehat{\rho}_T - \rho = \mathrm{O}_{\mathbb{P}}(T^{-1})$.*

Proof. We will prove only for $\psi_1 > 0$ and the estimate defined in (4.6.1) – as for Theorem 4.5.1, the other cases are completely analogous. Let us introduce the notation $\widehat{\tau}_T := \widehat{\rho}_T T$. We need to show

$$\lim_{K\to\infty} \sup_{T\in\mathbb{R}} \mathbb{P}(|\widehat{\tau}_T - \rho T| \geqslant K) = 0,$$

or, equivalently,

$$\lim_{K\to\infty} \limsup_{T\in\mathbb{R}} \mathbb{P}(|\widehat{\tau}_T - \rho T| \geqslant K) = 0.$$

For this, it is sufficient to show that

$$\lim_{K\to\infty} \limsup_{T\to\infty} \mathbb{P}\left(\sup_{\rho T - K < t < \rho T + K} \widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)} \leqslant \sup_{0 \leqslant t \leqslant \rho T - K} \widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)} \right) = 0 \tag{4.6.2}$$

and that

$$\lim_{K\to\infty} \limsup_{T\to\infty} \mathbb{P}\left(\sup_{\rho T - K < t < \rho T + K} \widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)} \leqslant \sup_{\rho T + K \leqslant t \leqslant T} \widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)} \right) = 0. \tag{4.6.3}$$

First we prove (4.6.2). We observe

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\rho T - K < t < \rho T + K} \widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)} \leqslant \sup_{0 \leqslant t \leqslant \rho T - K} \widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)} \right) \leqslant \mathbb{P}\left(\widehat{\boldsymbol{\mathcal{M}}}_{\rho T,1}^{(T)} \leqslant \sup_{0 \leqslant t \leqslant \rho T - K} \widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)} \right) \\
&= \mathbb{P}\left(\inf_{0 \leqslant t \leqslant \rho T - K} (\widehat{\boldsymbol{\mathcal{M}}}_{\rho T,1}^{(T)} - \widehat{\boldsymbol{\mathcal{M}}}_{t,1}^{(T)}) \leqslant 0 \right) = \mathbb{P}\left(\inf_{K \leqslant t \leqslant \rho T} T^{1/2} t^{-1} \int_{\rho T - t}^{\rho T} 1 \, \mathrm{d}\widehat{\boldsymbol{\mathcal{M}}}_{s,1}^{(T)} \leqslant 0 \right).
\end{aligned}$$

We apply the decomposition (4.5.1) and (4.5.2) to obtain

(4.6.4)

$$
\begin{aligned}
&\mathbb{P}\left(\inf_{K\leqslant t\leqslant \rho T} T^{1/2}t^{-1}\int_{\rho T-t}^{\rho T} 1\,\mathrm{d}\widehat{\boldsymbol{\mathcal{M}}}_{s,1}^{(T)} \leqslant 0\right)\\
&\quad\leqslant \mathbb{P}\left(\inf_{K\leqslant t\leqslant \rho T} t^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_{\rho T-t}^{\rho T}\begin{bmatrix}1 & -\mathbb{E}(Y_s)\\ -\mathbb{E}(Y_s) & \mathbb{E}(Y_s^2)\end{bmatrix}(\boldsymbol{\theta}'-\widetilde{\boldsymbol{\theta}})\,\mathrm{d}s \leqslant \frac{\psi_1}{2\rho}\right)\\
&\qquad+\mathbb{P}\left(\sup_{K\leqslant t\leqslant \rho T}\left|t^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_{\rho T-t}^{\rho T}\begin{bmatrix}1\\ -Y_s\end{bmatrix}\mathrm{d}M_s'\right| \geqslant \frac{\psi_1}{8\rho}\right)\\
&\qquad+\mathbb{P}\left(\sup_{K\leqslant t\leqslant \rho T}\left|t^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_{\rho T-t}^{\rho T}\begin{bmatrix}0 & \mathbb{E}(Y_s)-Y_s\\ \mathbb{E}(Y_s)-Y_s & Y_s^2-\mathbb{E}(Y_s^2)\end{bmatrix}(\boldsymbol{\theta}'-\widehat{\boldsymbol{\theta}}_T)\,\mathrm{d}s\right| \geqslant \frac{\psi_1}{8\rho}\right)\\
&\qquad+\mathbb{P}\left(\sup_{K\leqslant t\leqslant \rho T}\left|t^{-1}\mathbf{1}_1\widetilde{\boldsymbol{I}}^{-1/2}\int_{\rho T-t}^{\rho T}\begin{bmatrix}1 & -\mathbb{E}(Y_s)\\ -\mathbb{E}(Y_s) & \mathbb{E}(Y_s^2)\end{bmatrix}(\widetilde{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_T)\,\mathrm{d}s\right| \geqslant \frac{\psi_1}{8\rho}\right)\\
&\qquad+\mathbb{P}\left(\sup_{K\leqslant t\leqslant \rho T}\left|t^{-1}\mathbf{1}_1\left(\left(\frac{\boldsymbol{I}_T}{T}\right)^{-1/2}-\widetilde{\boldsymbol{I}}^{-1/2}\right)\int_{\rho T-t}^{\rho T}\begin{bmatrix}1\\ -Y_s\end{bmatrix}\mathrm{d}\widehat{M}_s^{(T)}\right| \geqslant \frac{\psi_1}{8\rho}\right).
\end{aligned}
$$

In the first term we take the probability of a deterministic event, therefore it is either 0 or 1; we show that for sufficiently large K, N it is 0. Actually, this is the same statement in continuous time as Lemma 2.9.5, and the proof is also essentially the same.

The fourth term in (4.6.4) converges to zero as $T\to\infty$ for any K. Indeed, we have

$$
\sup_{0\leqslant t\leqslant \rho T}\left\|t^{-1}\int_{\rho T-t}^{\rho T}\begin{bmatrix}1 & -\mathbb{E}(Y_s)\\ -\mathbb{E}(Y_s) & \mathbb{E}(Y_s^2)\end{bmatrix}\mathrm{d}s\right\| \leqslant \sup_{0\leqslant t\leqslant \rho T}\left\|\begin{bmatrix}1 & -\mathbb{E}(Y_t)\\ -\mathbb{E}(Y_t) & \mathbb{E}(Y_t^2)\end{bmatrix}\right\|,
$$

where the right hand side is bounded as $T\to\infty$, and $\widetilde{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_T\to 0$ a.s., which is sufficient. For the third term in (4.6.4) we use Lemma 4.7.6 and for the second one we can use Lemma 4.7.7. The only term that remains in (4.6.4) is the last one. This can be handled by the same method that we applied at the end of the proof of Theorem 4.5.1. Let us consider, again with the L_1-norm $\|\cdot\|$ and its induced matrix norm $\|\cdot\|_*$,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{K\leqslant t\leqslant \rho T}\left|t^{-1}\mathbf{1}_1\left(\left(\frac{\boldsymbol{I}_T}{T}\right)^{-1/2}-\widetilde{\boldsymbol{I}}^{-1/2}\right)\int_{\rho T-t}^{\rho T}\begin{bmatrix}1\\ -Y_s\end{bmatrix}\mathrm{d}\widehat{M}_s^{(T)}\right| \geqslant \frac{\psi_1}{8\rho}\right)\\
&\quad\leqslant \mathbb{P}\left(\left\|\left(\frac{\boldsymbol{I}_T}{T}\right)^{-1/2}-\widetilde{\boldsymbol{I}}^{-1/2}\right\|_* \geqslant \frac{\psi_1}{24\phi\rho}\right)+\mathbb{P}\left(\sup_{K\leqslant t\leqslant \rho T}\left\|t^{-1}\int_{\rho T-t}^{\rho T}\begin{bmatrix}1\\ -Y_s\end{bmatrix}\mathrm{d}\widehat{M}_s^{(T)}\right\| \geqslant 3\phi\right),
\end{aligned}
$$

with

$$\phi := \left\| \begin{bmatrix} 1 & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix} (\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}}) \right\|.$$

Taking the limit as $T \to \infty$ and then $K \to \infty$, the first term does not depend on K and tends to zero as $T \to \infty$, and for the second term we can use the same reasoning as in (4.6.4):

$$\begin{aligned}
&\mathbb{P}\left(\sup_{K \leqslant t \leqslant \rho T} \left\| t^{-1} \int_{\rho T - t}^{\rho T} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}\widehat{M}_s^{(T)} \right\| \geqslant 3\phi \right) \\
&\quad \leqslant \mathbb{P}\left(\sup_{K \leqslant t \leqslant \rho T} \left\| t^{-1} \int_{\rho T - t}^{\rho T} \begin{bmatrix} 1 & -\mathbb{E}(Y_s) \\ -\mathbb{E}(Y_s) & \mathbb{E}(Y_s^2) \end{bmatrix} (\boldsymbol{\theta}' - \widetilde{\boldsymbol{\theta}}) \, \mathrm{d}s \right\| \geqslant 2\phi \right) \\
&\quad + \mathbb{P}\left(\sup_{K \leqslant t \leqslant \rho T} \left\| t^{-1} \int_{\rho T - t}^{\rho T} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}M_s' \right\| \geqslant \frac{\phi}{3} \right) \\
&\quad + \mathbb{P}\left(\sup_{K \leqslant t \leqslant \rho T} \left\| t^{-1} \int_{\rho T - t}^{\rho T} \begin{bmatrix} 0 & \mathbb{E}(Y_s) - Y_s \\ \mathbb{E}(Y_s) - Y_s & Y_s^2 - \mathbb{E}(Y_s^2) \end{bmatrix} (\boldsymbol{\theta}' - \widehat{\boldsymbol{\theta}}_T) \, \mathrm{d}s \right\| \geqslant \frac{\phi}{3} \right) \\
&\quad + \mathbb{P}\left(\sup_{K \leqslant t \leqslant \rho T} \left\| t^{-1} \int_{\rho T - t}^{\rho T} \begin{bmatrix} 1 & -\mathbb{E}(Y_s) \\ -\mathbb{E}(Y_s) & \mathbb{E}(Y_s^2) \end{bmatrix} (\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_T) \, \mathrm{d}s \right\| \geqslant \frac{\phi}{3} \right).
\end{aligned}$$

The probability of the first term tends to zero, and the rest can be handled in exactly the same way as the corresponding terms in (4.6.4). For the proof of (4.6.3) we employ the same technique, but we need Lemmas 4.7.8 and 4.7.9 in place of 4.7.6 and 4.7.7. □

4.6.2 Remark. Again, this proof is in close parallel with the proof of Theorem 2.8.1. However, the underlying lemmas, which will be detailed in section 4.7, while structurally similar, are much more difficult to prove – especially Lemma 4.7.7, where we need to use an idea from the standard proof of the law of the iterated logarithm, and combine it with the Hájek–Rényi type inequality from Lemma 2.9.1.

4.7 Details of the proofs

In this section we detail the necessary lemmata for the proofs of our main theorems. Some of them, especially Lemma 4.7.1, are rather technical and depend essentially on tedious but straightforward calculations. Others, while using more sophisticated tools, are also tailored to the specific needs of the proofs and their proofs are not particularly insightful themselves, hence they were relegated to this section. The one exception to this is Lemma 4.7.2, which is an analogue of Lemma 2.9.1 and may deserve independent interest.

4.7.1 Lemma. *For a subcritical CIR process we have*

$$\text{(4.7.1)} \quad \operatorname{Var}\left(\int_0^t Y_s \mathrm{d}s\right) = \mathrm{O}(t), \quad t \to \infty, \qquad \textit{and} \qquad \operatorname{Var}\left(\int_0^t Y_s^2 \mathrm{d}s\right) = \mathrm{O}(t), \quad t \to \infty.$$

Proof. We show the first convergence only. We note

$$\operatorname{Var}\left(\int_0^t Y_s \mathrm{d}s\right) = \mathbb{E}\left(\int_0^t (Y_u - \mathbb{E}\, Y_u)\mathrm{d}u \int_0^t (Y_v - \mathbb{E}\, Y_v)\mathrm{d}v\right) = \iint\limits_{[0,t]^2} \operatorname{Cov}(Y_u, Y_v)\mathrm{d}u\mathrm{d}v$$

and similarly for $\operatorname{Var}\left(\int_0^t Y_s^2 \mathrm{d}s\right)$. From here, the proof could be finished simply by referring to Overbeck and Rydén (1997, (B.3) and (B.5)). However, we have chosen to detail our calculations (at least for $\operatorname{Cov}(Y_u, Y_v)$), because we will require the details later on. By using (3.1.1), we can write

$$\text{(4.7.2)} \quad \begin{aligned} \operatorname{Cov}(Y_u, Y_v) &= e^{-b(u+v)} \operatorname{Var}(Y_0) + \sigma^2 \int_0^{u \wedge v} e^{-b(u+v-2w)}\, \mathbb{E}(Y_w)\mathrm{d}w \\ &\leqslant e^{-b(u+v)} \operatorname{Var}(Y_0) + \left(E(Y_0) + \frac{a}{b}\right)\sigma^2 \int_0^{u \wedge v} e^{-b(u+v-2w)}\mathrm{d}w, \end{aligned}$$

since

$$\mathbb{E}(Y_w) = e^{-bw}\, \mathbb{E}(Y_0) + a \int_0^w e^{-bs}\mathrm{d}s.$$

Furthermore,

$$\text{(4.7.3)} \quad \begin{aligned} \iint\limits_{[0,t]^2} \left(\int_0^{u \wedge v} e^{-b(u+v-2w)}\mathrm{d}w\right)\mathrm{d}u\mathrm{d}v &= \iint\limits_{[0,t]^2} \left[\frac{1}{2b}\left(e^{-b|u-v|} - e^{-b(u+v)}\right)\right]\mathrm{d}u\mathrm{d}v \\ &\leqslant \frac{1}{b} \iint\limits_{[0,t]^2} e^{-b|u-v|}\mathrm{d}u\mathrm{d}v = \mathrm{O}(t). \end{aligned}$$

Recalling the last line of (4.7.2) and noting

$$\iint\limits_{[0,t]^2} e^{-b(u+v)}\mathrm{d}u\mathrm{d}v = \mathrm{O}(t),$$

the proof is complete. □

The following lemma is an analogue of Lemma 2.9.1, which is a Hájek–Rényi type inequality. With Lemma 2.9.1 one can estimate the tail probabilities of the maximum of a random sequence, based solely on the joint moments of the elements and, critically, without the assumption of independence. In our applications, not the supremum of a sequence but the maximum of a function is considered, so we had to modify the statement accordingly.

It turns out that the proof can be constructed along the lines of in Kokoszka and Leipus (2000, Theorem 4.1). In that paper, a slightly stronger result than Lemma 2.9.1 was formulated and proven; however, it was impractical to use, hence the more useful corollary formulated in Kokoszka and Leipus (1998, Theorem 3.1), which is obtainable from Kokoszka and Leipus (2000, Theorem 4.1) by a simple application of the Cauchy–Schwarz theorem.

4.7.2 Lemma. *Let Y_t be a process with a.s. continuous trajectory, $\alpha, \beta \in \mathbb{R}_+$ with $\alpha < \beta$ and c a deterministic function. Then, for any $\varepsilon > 0$,*

$$\varepsilon^2\, \mathbb{P}\left\{\sup_{s\in[\alpha,\beta]}\left(c(s)\int_0^s Y_u \mathrm{d}u\right)^2 > \varepsilon^2\right\} \leqslant c(\alpha)^2 \int_0^\alpha \mathbb{E}(Y_u^2)\,\mathrm{d}u$$
$$+\int_\alpha^\beta \left(\int_0^s\int_0^s \mathbb{E}(Y_uY_v)\mathrm{d}u\mathrm{d}v\right)\mathrm{d}|c(s)^2| + 2\int_\alpha^\beta c(s)^2\left[\mathbb{E}(Y_s^2)\int_0^s\int_0^s \mathbb{E}(Y_uY_v)\mathrm{d}u\mathrm{d}v\right]^{1/2}\mathrm{d}s$$

Proof. For any nonnegative process Z_t with a.s. continuous trajectories and a.s. locally bounded variation, let τ_ε be the first hitting time of $[\varepsilon,\infty)$ in $[\alpha,\infty)$, A be the event $\{\tau_\varepsilon < \beta\}$ and D_s be the event $\{\sup_{\alpha\leqslant u\leqslant s} Z_u \leqslant \varepsilon\}$. Note that $D_\beta = A^C$. Then it is easy to check that

$$\varepsilon \mathbf{1}_A \leqslant Z_\alpha + \int_\alpha^\beta \mathbf{1}_{D_s}\mathrm{d}Z_s.$$

Indeed, if A occurs, the LHS is ε, and the RHS is ε, if $Z_\alpha < \varepsilon$ and Z_α if $Z_\alpha \geqslant \varepsilon$. If A^C occurs, the LHS is zero, while the RHS is $Z_\beta \geqslant 0$.

Let us apply this result with $Z_t = c(t)^2\left|\int_0^t Y_s\,\mathrm{d}s\right|^2$ and for simplification let us introduce $K_\alpha := c(\alpha)^2\int_0^\alpha \mathbb{E}(Y_u^2)\,\mathrm{d}u$. We take expectations on both sides:

$$\varepsilon^2\,\mathbb{P}\left(\sup_{\alpha\leqslant s\leqslant\beta}\left|c(s)\int_0^s Y_u\mathrm{d}u\right| > \varepsilon\right) \leqslant \mathbb{E}\left[c(\alpha)^2\int_0^\alpha Y_u^2\,\mathrm{d}u\right] + \mathbb{E}\left[\int_\alpha^\beta \mathbf{1}_{D_s}\mathrm{d}\left(\left(c(s)\int_0^s Y_u\mathrm{d}u\right)^2\right)\right]$$
$$= K(\alpha) + \mathbb{E}\left[2\int_\alpha^\beta \mathbf{1}_{D_s}c(s)\int_0^s Y_u\mathrm{d}u\left(\left(\int_0^s Y_u\mathrm{d}u\right)\mathrm{d}c(s) + c(s)Y_s\mathrm{d}s\right)\right]$$
$$= K(\alpha) + \mathbb{E}\left[2\int_\alpha^\beta \mathbf{1}_{D_s}\left(\int_0^s\int_0^s Y_uY_v\mathrm{d}u\mathrm{d}v\right)\mathrm{d}(c^2(s)) + 2\int_\alpha^\beta \mathbf{1}_{D_s}c^2(s)Y_s\int_0^s Y_u\mathrm{d}u\mathrm{d}s\right]$$
$$\leqslant K(\alpha) + \mathbb{E}\left[2\int_\alpha^\beta \mathbf{1}_{D_s}\left(\int_0^s\int_0^s Y_uY_v\mathrm{d}u\mathrm{d}v\right)\mathrm{d}|c^2(s)| + 2\int_\alpha^\beta \mathbf{1}_{D_s}c^2(s)Y_s\int_0^s Y_u\mathrm{d}u\mathrm{d}s\right]$$

In the last step we replaced the induced norm of $c^2(s)$ by its total variation norm. Indeed, the inequality holds because $\int_0^s\int_0^s Y_uY_v\mathrm{d}u\mathrm{d}v = \left(\int_0^s Y_u\mathrm{d}u\right)^2$ for every ω in the probability space where Y is defined, therefore the integrand is nonnegative. Next, we replace the indicator function by one (this will not cause the expectation to decrease, since all integrands are nonnegative), and we employ several well-known inequalities to obtain our statement. □

4.7.3 Lemma. *If the parameters a and b remain constant, we have, for any $\gamma < \frac{1}{4}$,*

$$T^{\gamma-1} \sup_{0 \leqslant t \leqslant T} \int_0^t |Y_u - \mathbb{E}(Y_u)| \, \mathrm{d}u = \mathrm{o}_{\mathbb{P}}(1) \qquad \textit{and} \qquad T^{\gamma-1} \sup_{0 \leqslant t \leqslant T} \int_0^t |Y_u^2 - \mathbb{E}(Y_u^2)| \, \mathrm{d}u = \mathrm{o}_{\mathbb{P}}(1).$$

Proof. We show that the suprema are $\mathrm{O}_{\mathbb{P}}(1)$, but this is the same statement, since for any given γ we can take $\frac{1}{2} > \gamma' > \gamma$, apply the lemma with γ' and then multiply the suprema with $T^{\gamma-\gamma'}$, which is clearly $\mathrm{o}(1)$. Also, we only show the proof of the first statement; the proof of the second one is completely analogous. What we actually prove is the slightly stronger statement

$$\sup_{0 \leqslant t \leqslant T} t^{\gamma-1} \int_0^t |Y_u - \mathbb{E}(Y_u)| \, \mathrm{d}u = \mathrm{O}_{\mathbb{P}}(1).$$

We will use Lemma 4.7.2 for the process $Y_t := Y_t - \mathbb{E}(Y_t)$ and $c(s) = s^{\gamma-1}$ and $\alpha = 0, \ \beta = T$. Then we can use Lemma 4.7.1 to conclude that

$$\int_0^s \int_0^s \mathbb{E}(Y_u Y_v) \, \mathrm{d}v \, \mathrm{d}u = \int_0^s \int_0^s \mathrm{Cov}(Y_u, Y_v) \, \mathrm{d}v \, \mathrm{d}u \leqslant \kappa s, \quad s \in \mathbb{R}_+,$$

for some constant $\kappa > 0$. Hence, in this case,

$$\int_0^T \left(\int_0^s \int_0^s \mathbb{E}(Y_u Y_v) \mathrm{d}u \mathrm{d}v \right) \mathrm{d}|c(s)^2| + 2 \int_0^T c(s)^2 \left[\mathbb{E}(Y_s^2) \int_0^s \int_0^s \mathbb{E}(Y_u Y_v) \mathrm{d}u \mathrm{d}v \right]^{1/2} \mathrm{d}s$$
$$\leqslant \kappa(2 - 2\gamma) \int_0^T s^{2\gamma-2} \, \mathrm{d}s + 2(K\kappa)^{1/2} \int_0^T s^{2\gamma-3/2} \, \mathrm{d}s < \infty.$$

This implies the desired statement immediately. $\square$

4.7.4 Lemma. *If the parameters a and b remain constant, we have, for any $\gamma < \frac{1}{2}$,*

$$\sup_{0 \leqslant t \leqslant T} T^{\gamma-1} |M_t| = \mathrm{o}_{\mathbb{P}}(1) \qquad \textit{and} \qquad \sup_{0 \leqslant t \leqslant T} T^{\gamma-1} \left| \int_0^t Y_s \mathrm{d}M_s \right| = \mathrm{o}_{\mathbb{P}}(1).$$

Proof. Similarly to the previous lemma, we only show $\mathrm{O}_{\mathbb{P}}(1)$ for the first statement. First we note that $(M_t)_{t \in \mathbb{R}_+}$ has an a.s. continuous trajectory on $\mathbb{R}_+$, therefore also on $[0, 1]$. Thus we conclude that $\sup_{0 \leqslant t \leqslant 1} |M_t| = \mathrm{O}_{\mathbb{P}}(1)$. Next, we use the law of the iterated logarithm for continuous martingales. This can be put together from the Dambis–Dubins–Schwarz theorem (Karatzas and Shreve, 1991, Theorem 3.4.6) and the law of the iterated logarithm for the Wiener process (Karatzas and Shreve, 1991, Theorem 2.9.23).

$$\limsup_{t \to \infty} \frac{|M_t|}{\sigma^{2\lambda} \left(\int_0^t Y_u \, \mathrm{d}u \right)^{\lambda}} \leqslant \limsup_{t \to \infty} \frac{|M_t|}{\sigma \sqrt{\int_0^t Y_u \, \mathrm{d}u} \sqrt{\log \log(\sigma^2 \int_0^t Y_u \, \mathrm{d}u)}} = 1 \quad \text{a.s.}, \quad \forall \lambda > \frac{1}{2},$$

which means that the supremum on $[1, \infty]$ is finite a.s. (since the process in question has a.s. continuous trajectories).

Next we note that
$$\frac{\sigma^{2\lambda}\left(\int_0^t Y_u\,\mathrm{d}u\right)^{\lambda}}{t^{\lambda}} \to \sigma^{2\lambda}\,\mathbb{E}(Y_\infty)^{\lambda} \quad \text{a.s..}$$
Now the statement of the lemma is obtained straightforwardly since
$$\sup_{0\leqslant t\leqslant T} T^{\gamma-1}|M_t| = \max\left(\sup_{0\leqslant t\leqslant 1} T^{\gamma-1}|M_t|, \sup_{1\leqslant t\leqslant T} T^{\gamma-1}|M_t|\right),$$
and both terms have been shown to be $\mathrm{O}_{\mathbb{P}}(1)$. $\square$

4.7.5 Lemma. *Under the conditions of Theorem 4.5.1,*
$$\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}} = \mathrm{O}_{\mathbb{P}}(T^{-1/2}).$$

Proof. We have
$$T^{1/2}(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}) = (T^{-1}\boldsymbol{Q}_T)^{-1}T^{-1/2}\left[\boldsymbol{d}_{0,\tau} - \boldsymbol{Q}_T\widetilde{\boldsymbol{Q}}^{-1}\left(\rho\boldsymbol{Q}'\begin{bmatrix} a \\ b\end{bmatrix}\right)\right.$$
$$\left. + \boldsymbol{d}_{\tau,T} - \boldsymbol{Q}_T\widetilde{\boldsymbol{Q}}^{-1}\left((1-\rho)\boldsymbol{Q}''\begin{bmatrix} a'' \\ b''\end{bmatrix}\right)\right].$$

The first factor converges almost surely, so we analyze
$$T^{-1/2}\left[\boldsymbol{d}_{0,\tau} - \boldsymbol{Q}_T\widetilde{\boldsymbol{Q}}^{-1}\left(\rho\boldsymbol{Q}'\begin{bmatrix} a \\ b\end{bmatrix}\right)\right] = T^{-1/2}\widetilde{\boldsymbol{d}}_\tau + T^{-1/2}\left(\boldsymbol{Q}_{[0,\tau]} - \boldsymbol{Q}_T\widetilde{\boldsymbol{Q}}^{-1}\rho\boldsymbol{Q}'\right)\begin{bmatrix} a' \\ b'\end{bmatrix}.$$
The first term is $\mathrm{O}_{\mathbb{P}}(1)$ by (4.2.6). We need to show that the second term is also $\mathrm{O}_{\mathbb{P}}(1)$. For this, we can neglect the vector of the parameters, which are constant, so we investigate
$$\begin{aligned} T^{-1/2}\left(\boldsymbol{Q}_\tau - \rho\boldsymbol{Q}_T\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{Q}'\right) &= T^{-1/2}\left(\boldsymbol{Q}_\tau - \mathbb{E}(\boldsymbol{Q}_\tau)\right) + T^{-1/2}\left(\mathbb{E}(\boldsymbol{Q}_\tau) - \tau\boldsymbol{Q}'\right) \\ &\quad - T^{-1/2}\left(\rho(\boldsymbol{Q}_T - \mathbb{E}(\boldsymbol{Q}_T))\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{Q}'\right) \\ &\quad - T^{-1/2}\left(\rho(\mathbb{E}(\boldsymbol{Q}_T) - T\widetilde{\boldsymbol{Q}})\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{Q}'\right). \end{aligned}$$
The first and third factors have a finite variance at the limit, by Lemma 4.7.1. Therefore, by an application of Chebyshev's inequality, we have that they are $\mathrm{O}_{\mathbb{P}}(1)$. The second and fourth terms are deterministic and $\mathrm{O}(1)$ by (3.2.7). $\square$

4.7.6 Lemma. *Under the conditions of Theorem 4.5.1 we have, for an arbitrary* $\varepsilon > 0$,
$$\lim_{K\to\infty}\limsup_{T\to\infty}\mathbb{P}\left(\sup_{K\leqslant t\leqslant\rho T}\left|t^{-1}\int_{\rho T-t}^{\rho T}(Y_s - \mathbb{E}(Y_s))\,\mathrm{d}s\right| > \varepsilon\right) = 0.$$

and

$$\lim_{K\to\infty}\limsup_{T\to\infty}\mathbb{P}\left(\sup_{K\leqslant t\leqslant\rho T}\left|t^{-1}\int_{\rho T-t}^{\rho T}(Y_s^2-\mathbb{E}(Y_s^2))\,\mathrm{d}s\right|>\varepsilon\right)=0.$$

Proof. As for the previous lemmas, we only prove the first statement. We use Lemma 4.7.2. We choose $c(s)=s^{-1}$ and $Y_s=Y_{\rho T-s}-\mathbb{E}(Y_{\rho T-s})$ with $\alpha=K$ and $\beta=\rho T$. The estimate on the probability in question is then

$$\begin{aligned}K^{-2}\int_{\rho T-K}^{\rho T}\operatorname{Var}(Y_u)\,\mathrm{d}u+\int_K^{\rho T}\left(\int_{\rho T-s}^{\rho T}\int_{\rho T-s}^{\rho T}\operatorname{Cov}(Y_u,Y_v)\mathrm{d}u\mathrm{d}v\right)\mathrm{d}\left|s^{-2}\right|\\+2\int_K^{\rho T}s^{-2}\left[\operatorname{Var}(Y_s)\int_{\rho T-s}^{\rho T}\int_{\rho T-s}^{\rho T}\operatorname{Cov}(Y_u,Y_v)\mathrm{d}u\mathrm{d}v\right]^{1/2}\mathrm{d}s.\end{aligned}\tag{4.7.4}$$

Now we make use of (4.7.2) and (4.7.3) to show that

$$\begin{aligned}\int_{\rho T-s}^{\rho T}\int_{\rho T-s}^{\rho T}\operatorname{Cov}(Y_u,Y_v)\mathrm{d}u\mathrm{d}v\leqslant\operatorname{Var}(Y_0)\int_{\rho T-s}^{\rho T}\int_{\rho T-s}^{\rho T}\mathrm{e}^{-b(u+v)}\mathrm{d}u\mathrm{d}v\\+(\mathbb{E}(Y_0)+ab^{-1})\sigma^2b^{-1}\int_{\rho T-s}^{\rho T}\int_{\rho T-s}^{\rho T}\mathrm{e}^{-b|u-v|}\mathrm{d}u\mathrm{d}v\leqslant\mu s,\end{aligned}$$

for some positive constant μ. We introduce $\lambda:=\sup_{t\in\mathbb{R}}\operatorname{Var}(Y_t)<\infty$, to continue the estimation started in (4.7.4):

$$K^{-2}K\lambda+2\int_K^{\rho}Ts^{-3}\mu s\,\mathrm{d}s+2\int_K^{\rho T}s^{-2}(\lambda\mu)^{1/2}s^{1/2}\,\mathrm{d}s.$$

Clearly, as $T\to\infty$ (and hence $\rho T\to\infty$), and then $K\to\infty$, this expression tends to zero, which completes our proof. □

For the next lemma we will need to recall Lemma 2.9.1 once more.

4.7.7 Lemma. *Under the conditions of Theorem 4.5.1 we have, for any $\varepsilon>0$,*

$$\lim_{K\to\infty}\limsup_{T\to\infty}\mathbb{P}\left(\sup_{K\leqslant t\leqslant\rho T}\left|t^{-1}(M'_{\rho T}-M'_{\rho T-t})\right|>\varepsilon\right)=0.$$

and

$$\lim_{K\to\infty}\limsup_{T\to\infty}\mathbb{P}\left(\sup_{K\leqslant t\leqslant\rho T}\left|t^{-1}\left(\int_{\rho T-t}^{\rho T}Y_s\mathrm{d}M'_s\right)\right|>\varepsilon\right)=0.$$

Proof. We prove the first statement first. Let us take a backward partition of $[0,\rho T]$ such that $0=t_n<t_{n-1}<t_{n-2}<\ldots<t_1<t_0=\rho T$. For $t\in[t_{i+1},t_i]$, we have

$$\left|\frac{M'_{\rho T}-M'_t}{\rho T-t}\right|\leqslant\left|\frac{M'_{\rho T}-M'_{t_{i+1}}}{\rho T-t_i}\right|+\left|\frac{M'_t-M'_{t_{i+1}}}{\rho T-t_i}\right|.$$

Therefore, we have the following estimation:

(4.7.5)
$$\begin{aligned}
&\mathbb{P}\left(\sup_{K\leqslant t\leqslant \rho T}\left|t^{-1}(M'_\rho T - M'_{\rho T-t})\right| > \varepsilon\right) = \mathbb{P}\left(\sup_{0\leqslant t\leqslant \rho T-K}\left|(\rho T-t)^{-1}(M'_{\rho T} - M'_t)\right| > \varepsilon\right)\\
&\leqslant \mathbb{P}\left(\max_{i^*\leqslant i\leqslant n}\left|(\rho T-t_i)^{-1}(M'_{\rho T} - M'_{t_{i+1}})\right| > \frac{\varepsilon}{2}\right)\\
&\quad + \sum_{i=i*}^{n}\mathbb{P}\left(\sup_{t_{i+1}<t<t_i}\left|(\rho T-t_i)^{-1}(M'_t - M'_{t_{i+1}})\right| > \frac{\varepsilon}{2}\right),
\end{aligned}$$

where $i^* = \min\{i : t_i < \rho T - K\}$. Let us use this estimate with $t_i := \rho T - 2^{i-1}$ for $0 < i < n$, so that $n = \lfloor \log_2 \rho T\rfloor$ and $i^* = \lfloor \log_2 K\rfloor + 1$.

Let us now apply Lemma 2.9.1 with $Y_1 := M'_{t_{i^*}} - M'_{\rho T}$, and $Y_k = M'_{t_{i^*+k-1}} - M'_{t_{i^*+k-2}}$ for $1 < k \leqslant n - i^* + 1$ and $c_k = (\rho T - t_{i^*+k-1})^{-1}$. Let us note that due to the structure of the t_i, we have $c_k = 2^{-(i^*+k-2)}$ for $k \leqslant n - i^*$ and $2^{-(n-1)} < c_{n-i^*+1} < c_{n-i^*}$. Consequently, we can use $|c_{k+1}^2 - c_k^2| \leqslant |c_{k+1}^2 - 4c_{k+1}^2| = 3c_{k+1}^2$. Also, notice that

$$\begin{aligned}
\sum_{i,j=1}^{k}\mathbb{E}(Y_iY_j) = \mathbb{E}\left(\sum_{i=1}^{k} Y_i\right)^2 &= \mathbb{E}(M'_{t_{i^*+k-1}} - M'_{\rho T})^2\\
&= \sigma^2\int_{t_{i^*+k-1}}^{\rho T}\mathbb{E}(Y_u)\,\mathrm{d}u \leqslant \sigma^2\mu(\rho T - t_{i^*+k-1}),
\end{aligned}$$

with $\mu = \sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t) < \infty$, and that, similarly,

$$\mathbb{E}(Y_{k+1}^2) \leqslant \sigma^2\mu(t_{i^*+k} - t_{i^*+k-1}) = \sigma^2\mu 2^{(i^*+k-2)}.$$

All in all, with Lemma 2.9.1, we can estimate the first term in (4.7.5) by $\mu 2^{-i^*}$. This does not depend on n (hence, on T), and since $i^* \to \infty$ as $K \to \infty$, we have that the first term in (4.7.5) converges to zero as $\rho T \to \infty$ and then $K \to \infty$.

For the second term in (4.7.5) we will use Doob's submartingale inequality (see, e.g., Karatzas and Shreve, 1991, Theorem 1.3.8. (i)) to the submartingales

$$N_{t,i} := (M'_{t_{i+1}+t} - M'_{t_{i+1}})^2, \quad t\in[0, t_i - t_{i+1}], \quad i = i^*, \dots, n,$$

for which clearly

$$\mathbb{P}\left(\sup_{t_{i+1}<t<t_i}\left|(\rho T-t_i)^{-1}(M'_t - M'_{t_{i+1}})\right| > \frac{\varepsilon}{2}\right) = \mathbb{P}\left(\sup_{0\leqslant t\leqslant t_i-t_{i+1}} N_{t,i} > \frac{\varepsilon^2(\rho T - t_i)^2}{4}\right).$$

The inequality states that

$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant t_i-t_{i+1}} N_{t,i} > \frac{\varepsilon^2(\rho T-t_i)^2}{4}\right) \leqslant \frac{4\,\mathbb{E}(N_{t_i-t_{i+1}})}{\varepsilon^2(\rho T-t_i)^2} = \frac{4\,\mathbb{E}(M'_{t_i}-M'_{t_{i+1}})^2}{\varepsilon^2(\rho T-t_i)^2} \leqslant \frac{4\sigma^2\mu(t_i-t_{i+1})}{\varepsilon^2(\rho T-t_i)^2}.$$

Now, in our present setting, $t_i - t_{i+1} \leqslant 2^{i-1}$ and $(\rho T - t_i)^2 \geqslant 2^{2i-4}$. Thus, the second term in (4.7.5) can be estimated from above by

$$\frac{\sigma^2\mu\varepsilon^2}{4}\sum_{i=i^*}^{n} 2^{-i+3} \leqslant \frac{\sigma^2\mu\varepsilon^2}{4}2^{-i^*+3}\sum_{i=0}^{\infty}2^{-i}.$$

Again, clearly this does not depend on n (thus, T) and converges to zero as $i^* \to \infty$ (and thus, as $K \to \infty$). This suffices for the first statement. The second one can be obtained in a completely analogous way, since we only used the fact that $(M'_t)_{t\in[0,\infty]}$ is a martingale with essentially linear quadratic variation, which is also true of $\left(\int_0^t Y_s \mathrm{d}M'_s\right)_{t\in[0,\infty]}$. □

The following lemma is the "forward" analogue of Lemma 4.7.6, and its proof is also the same, with straightforward modifications.

4.7.8 Lemma. *Under the conditions of Theorem 4.5.1 we have, for an arbitrary $\varepsilon > 0$,*

$$\lim_{K\to\infty}\limsup_{T\to\infty}\mathbb{P}\left(\sup_{K\leqslant t\leqslant(1-\rho)T}\left|t^{-1}\int_{\rho T}^{\rho T+t}(Y_s-\mathbb{E}(Y_s))\,\mathrm{d}s\right| > \varepsilon\right) = 0.$$

and

$$\lim_{K\to\infty}\limsup_{T\to\infty}\mathbb{P}\left(\sup_{K\leqslant t\leqslant(1-\rho)T}\left|t^{-1}\int_{\rho T}^{\rho T+t}(Y_s^2-\mathbb{E}(Y_s^2))\,\mathrm{d}s\right| > \varepsilon\right) = 0.$$

The forward analogue of Lemma 4.7.7 can also be proved in the same manner as the original:

4.7.9 Lemma. *Under the conditions of Theorem 4.5.1 we have, for any $\varepsilon > 0$,*

$$\lim_{K\to\infty}\limsup_{T\to\infty}\mathbb{P}\left(\sup_{K\leqslant t\leqslant(1-\rho)T}\left|t^{-1}(M''_{\rho T+t}-M''_{\rho T})\right| > \varepsilon\right) = 0.$$

and

$$\lim_{K\to\infty}\limsup_{T\to\infty}\mathbb{P}\left(\sup_{K\leqslant t\leqslant(1-\rho)T}\left|t^{-1}\left(\int_{\rho T}^{\rho T+t} Y_s\mathrm{d}M''_s\right)\right| > \varepsilon\right) = 0.$$

Chapter 5

Estimates in the continuous case

In this chapter, we are going to derive CLS-like estimates for the Heston model from (3.0.1). As noted in section 1.4, we are going to transform the parameter space first, derive estimates for the transformed parameters, and estimate the original parameters by applying the inverse transformation.

5.1 Estimates for the transformed parameters

Using (3.1.4) and (3.1.5), by an easy calculation, for all $i \in \mathbb{N}$,

$$(5.1.1)\quad \mathbb{E}\left(\begin{bmatrix} Y_i \\ X_i \end{bmatrix} \,\middle|\, \mathcal{F}_{i-1}\right) = \begin{bmatrix} \mathrm{e}^{-b} & 0 \\ -\beta \int_0^1 \mathrm{e}^{-bu}\,\mathrm{d}u & 1 \end{bmatrix} \begin{bmatrix} Y_{i-1} \\ X_{i-1} \end{bmatrix} + \begin{bmatrix} \int_0^1 \mathrm{e}^{-bu}\,\mathrm{d}u & 0 \\ -\beta \int_0^1 \left(\int_0^u \mathrm{e}^{-bv}\,\mathrm{d}v\right)\mathrm{d}u & 1 \end{bmatrix} \begin{bmatrix} a \\ \alpha \end{bmatrix}.$$

Using that $\sigma(X_1, Y_1, \ldots, X_{i-1}, Y_{i-1}) \subseteq \mathcal{F}_{i-1}$, $i \in \mathbb{N}$, by the tower rule for conditional expectations, we have

$$\begin{aligned} \mathbb{E}\left(\begin{bmatrix} Y_i \\ X_i \end{bmatrix} \,\middle|\, \sigma(X_1, Y_1, \ldots, X_{i-1}, Y_{i-1})\right) &= \mathbb{E}\left(\mathbb{E}\left(\begin{bmatrix} Y_i \\ X_i \end{bmatrix} \,\middle|\, \mathcal{F}_{i-1}\right) \,\middle|\, \sigma(X_1, Y_1, \ldots, X_{i-1}, Y_{i-1})\right) \\ &= \begin{bmatrix} \mathrm{e}^{-b} & 0 \\ -\beta \int_0^1 \mathrm{e}^{-bu}\,\mathrm{d}u & 1 \end{bmatrix} \begin{bmatrix} Y_{i-1} \\ X_{i-1} \end{bmatrix} + \begin{bmatrix} \int_0^1 \mathrm{e}^{-bu}\,\mathrm{d}u & 0 \\ -\beta \int_0^1 \left(\int_0^u \mathrm{e}^{-bv}\,\mathrm{d}v\right)\mathrm{d}u & 1 \end{bmatrix} \begin{bmatrix} a \\ \alpha \end{bmatrix}, \qquad i \in \mathbb{N}, \end{aligned}$$

and hence a CLSE of (a, b, α, β) based on discrete time observations $(Y_i, X_i)_{i \in \{1,\ldots,n\}}$ could be obtained by solving the extremum problem

$$(5.1.2)\qquad \operatorname*{arg\,min}_{(a,b,\alpha,\beta)\in\mathbb{R}^4} \sum_{i=1}^{n} \left[(Y_i - dY_{i-1} - c)^2 + (X_i - X_{i-1} - \gamma - \delta Y_{i-1})^2\right],$$

where the transformed parameters are

(5.1.3)
$$
\begin{aligned}
&d := d(b) := \mathrm{e}^{-b}, \qquad c := c(a,b) := a \int_0^1 \mathrm{e}^{-bu}\,\mathrm{d}u,\\
&\delta := \delta(b,\beta) := -\beta \int_0^1 \mathrm{e}^{-bu}\,\mathrm{d}u, \qquad \gamma := \gamma(a,b,\alpha,\beta) := \alpha - a\beta \int_0^1 \left(\int_0^u \mathrm{e}^{-bv}\,\mathrm{d}v \right) \mathrm{d}u.
\end{aligned}
$$

Minimizing the right hand side with respect to $(c,d,\gamma,\delta) \in \mathbb{R}^4$ leads to

$$
\begin{bmatrix} \widehat{c}_n \\ \widehat{d}_n \\ \widehat{\gamma}_n \\ \widehat{\delta}_n \end{bmatrix} = \left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right) \right)^{-1} \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Y_i Y_{i-1} \\ X_n - x_0 \\ \sum_{i=1}^n (X_i - X_{i-1}) Y_{i-1} \end{bmatrix}, \tag{5.1.4}
$$

provided that $n \sum_{i=1}^n Y_{i-1}^2 > \left(\sum_{i=1}^n Y_{i-1} \right)^2$. Indeed, with the notation

$$
f(c,d,\gamma,\delta) := \sum_{i=1}^n \left[(Y_i - dY_{i-1} - c)^2 + (X_i - X_{i-1} - \gamma - \delta Y_{i-1})^2 \right], \qquad (c,d,\gamma,\delta) \in \mathbb{R}^4,
$$

we have

$$
\begin{aligned}
\frac{\partial f}{\partial c}(c,d,\gamma,\delta) &= -2 \sum_{i=1}^n (Y_i - dY_{i-1} - c),\\
\frac{\partial f}{\partial d}(c,d,\gamma,\delta) &= -2 \sum_{i=1}^n Y_{i-1}(Y_i - dY_{i-1} - c),\\
\frac{\partial f}{\partial \gamma}(c,d,\gamma,\delta) &= -2 \sum_{i=1}^n (X_i - X_{i-1} - \gamma - \delta Y_{i-1}),\\
\frac{\partial f}{\partial \delta}(c,d,\gamma,\delta) &= -2 \sum_{i=1}^n Y_{i-1}(X_i - X_{i-1} - \gamma - \delta Y_{i-1}).
\end{aligned}
$$

Hence the system of equations consisting of the first order partial derivatives of f being equal to 0 takes the form

$$
\left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right) \right) \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Y_{i-1} Y_i \\ X_n - x_0 \\ \sum_{i=1}^n (X_i - X_{i-1}) Y_{i-1} \end{bmatrix}.
$$

This implies (5.1.4), since the 4×4 matrix consisting of the second order partial derivatives of f having the form

$$2\left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top\right)\right)$$

is positive definite provided that $n\sum_{i=1}^n Y_{i-1}^2 > (\sum_{i=1}^n Y_{i-1})^2$. In fact, it turned out that for the calculation of the CLSE of (c, d, γ, δ), one does not need to know the values of the parameters σ_1, σ_2 and ϱ.

The next lemma assures the unique existence of the CLSE of (c, d, γ, δ) based on discrete time observations. Note that it is valid for all $b \in \mathbb{R}$, i.e., not only for the subcritical Heston model.

5.1.1 Lemma. *If $a \in \mathbb{R}_{++}$, $b \in \mathbb{R}$, $\sigma_1 \in \mathbb{R}_{++}$, and $Y_0 = y_0 \in \mathbb{R}_+$, then for all $n \geqslant 2$, $n \in \mathbb{N}$, we have*

$$\mathbb{P}\left(n\sum_{i=1}^n Y_{i-1}^2 > \left(\sum_{i=1}^n Y_{i-1}\right)^2\right) = 1,$$

and hence, supposing also that $\alpha, \beta \in \mathbb{R}$, $\sigma_2 \in \mathbb{R}_{++}$, $\varrho \in (-1, 1)$, there exists a unique CLSE $(\widehat{c}_n, \widehat{d}_n, \widehat{\gamma}_n, \widehat{\delta}_n)$ of (c, d, γ, δ) which has the form given in (5.1.4).

Proof. By an easy calculation,

$$n\sum_{i=1}^n Y_{i-1}^2 - \left(\sum_{i=1}^n Y_{i-1}\right)^2 = n\sum_{i=1}^n \left(Y_{i-1} - \frac{1}{n}\sum_{j=1}^n Y_{j-1}\right)^2 \geqslant 0,$$

and equality holds if and only if

$$Y_{i-1} = \frac{1}{n}\sum_{j=1}^n Y_{j-1}, \qquad i = 1, \ldots, n \qquad \Longleftrightarrow \qquad Y_0 = Y_1 = \cdots = Y_{n-1}.$$

Then, for all $n \geqslant 2$,

$$\mathbb{P}(Y_0 = Y_1 = \cdots = Y_{n-1}) \leqslant \mathbb{P}(Y_0 = Y_1) = \mathbb{P}(Y_1 = y_0) = 0,$$

since the law of Y_1 is absolutely continuous, see, e.g., Cox et al. (1985, formula 18). □

5.2 Asymptotic results for the transformed parameters

5.2.1 Theorem. *For a subcritical Heston model, that is, if $b \in \mathbb{R}_{++}$ and*

$$(Y_0, X_0) = (y_0, x_0) \in \mathbb{R}_{++} \times \mathbb{R},$$

the CLSE $(\widehat{c}_n, \widehat{d}_n, \widehat{\gamma}_n, \widehat{\delta}_n)$ *in* (5.1.4) *is strongly consistent and asymptotically normal, i.e.,*

$$(\widehat{c}_n, \widehat{d}_n, \widehat{\gamma}_n, \widehat{\delta}_n) \xrightarrow{\text{a.s.}} (c, d, \gamma, \delta) \qquad \text{as } n \to \infty,$$

and

$$\sqrt{n} \begin{bmatrix} \widehat{c}_n - c \\ \widehat{d}_n - d \\ \widehat{\gamma}_n - \gamma \\ \widehat{\delta}_n - \delta \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_4(\mathbf{0}, \boldsymbol{G}) \qquad \text{as } n \to \infty,$$

with

$$\boldsymbol{G} := \left(\boldsymbol{E}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix} \right)^{-1} \boldsymbol{D} \left(\boldsymbol{E}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix} \right)^{-1}, \tag{5.2.1}$$

where $\boldsymbol{D}$ *is defined in* (5.3.4)*. Furthermore,* $\boldsymbol{G}$ *is strictly positive definite.*

Proof. Due to Theorem 5.1.1 we will assume that the estimators exist uniquely in the form given by (5.1.4). Then we get

(5.2.2)

$$\begin{aligned}
\begin{bmatrix} \widehat{c}_n \\ \widehat{d}_n \\ \widehat{\gamma}_n \\ \widehat{\delta}_n \end{bmatrix}
&= \left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \right) \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Y_i Y_{i-1} \\ X_i - X_{i-1} \\ \sum_{i=1}^n (X_i - X_{i-1}) Y_{i-1} \end{bmatrix} \\
&= \left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \right) \left(\sum_{i=1}^n \begin{bmatrix} Y_i \\ X_i - X_{i-1} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \right) \\
&= \left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \right) \left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right) \right) \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix} \\
&\quad + \left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \right) \left(\sum_{i=1}^n \begin{bmatrix} Y_i - c - dY_{i-1} \\ X_i - X_{i-1} - \gamma - \delta Y_{i-1} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \right).
\end{aligned}$$

The final step depends on

$$\left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top\right)\right) \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix} = \left(\sum_{i=1}^n \boldsymbol{E}_2 \otimes \left(\begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top\right)\right) \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix}$$

$$= \left(\sum_{i=1}^n \begin{bmatrix} 1 & Y_{i-1} & 0 & 0 \\ Y_{i-1} & Y_{i-1}^2 & 0 & 0 \\ 0 & 0 & 1 & Y_{i-1} \\ 0 & 0 & Y_{i-1} & Y_{i-1}^2 \end{bmatrix}\right) \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix} = \sum_{i=1}^n \begin{bmatrix} c + dY_{i-1} \\ (c + dY_{i-1})Y_{i-1} \\ \gamma + \delta Y_{i-1} \\ (\gamma + \delta Y_{i-1})Y_{i-1} \end{bmatrix}.$$

Continuing from (5.2.2),

$$\begin{aligned} \begin{bmatrix} \widehat{c}_n \\ \widehat{d}_n \\ \widehat{\gamma}_n \\ \widehat{\delta}_n \end{bmatrix} &= \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix} + \left(\boldsymbol{E}_2 \otimes \left(\frac{1}{n}\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top\right)^{-1}\right) \left(\frac{1}{n}\sum_{i=1}^n \begin{bmatrix} \varepsilon_i \\ \eta_i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}\right) \\ &= \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix} + \left(\boldsymbol{E}_2 \otimes \left(\frac{1}{n}\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top\right)^{-1}\right) n^{-1/2}\boldsymbol{M}_{n,n}, \end{aligned} \tag{5.2.3}$$

where $\varepsilon_i := Y_i - c - dY_{i-1}$ and $\eta_i := X_i - X_{i-1} - \gamma - \delta Y_{i-1}$, also,

$$\boldsymbol{M}_{n,k} := n^{-\frac{1}{2}} \sum_{i=1}^k \begin{bmatrix} \varepsilon_i \\ \eta_i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}, \qquad n \in \mathbb{N}, \quad k \in \{1, \ldots, n\}. \tag{5.2.4}$$

The final thing to note is that, by (3.2.2) and (3.2.4),
(5.2.5)

$$\left(\frac{1}{n}\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top\right)^{-1} = \begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^n Y_{i-1} \\ \frac{1}{n}\sum_{i=1}^n Y_{i-1} & \frac{1}{n}\sum_{i=1}^n Y_{i-1}^2 \end{bmatrix}^{-1} \xrightarrow{\text{a.s.}} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}$$

where we used that

$$\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2 = \frac{a\sigma_1^2}{2b^2} \in \mathbb{R}_{++},$$

and consequently, the limit is indeed non-singular. The result is now a direct consequence of Slutsky's lemma and Lemmas 5.3.1, 5.3.2 and 5.3.3. □

5.2.2 Remark. The structure of (5.2.3) is essentially the same as (2.5.6). However, the resulting martingale is much more complicated to handle – indeed, the calculation of the quadratic variations and the condition checks of the martingale central limit theorem constitute a considerable part of the proof and indicates the hardships that are to be expected by carrying the research further in this direction, that is, trying to find the analogue of Theorem 2.5.1 in this setting.

5.3 Auxiliary lemmas

First we will check the conditions of the martingale central limit theorem in our setting. The reason for this is that the quadratic variation will be calculated as a by-product and we will require that for checking the conditions of the strong law of large numbers.

5.3.1 Lemma. *Under the conditions of Theorem 5.2.1 and $\boldsymbol{M}_{n,k}$ defined in* (5.2.4)*, we have*

$$\boldsymbol{M}_{n,n} = n^{-1/2} \sum_{k=1}^{n} \begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_4\left(\mathbf{0}, \boldsymbol{D}\right),$$

where $\boldsymbol{D}$ is defined in (5.3.4)*.*

Proof. We are going to apply the martingale central limit theorem (see Theorem 3.3.4) to $\boldsymbol{M}_{n,k}$ with the following choices: $d = 4$, $k_n = n$, $n \in \mathbb{N}$, $\mathcal{F}_{n,k} = \mathcal{F}_k$, $n \in \mathbb{N}$, $k \in \{1, \ldots, n\}$. First we check that our process is indeed a martingale. By (5.1.1) and (5.1.3),

$$\mathbb{E}(Y_i \,|\, \mathcal{F}_{i-1}) = dY_{i-1} + c, \quad i \in \mathbb{N},$$

and hence $(\varepsilon_i)_{i\in\mathbb{N}}$ is a sequence of martingale differences with respect to the filtration $(\mathcal{F}_i)_{i\in\mathbb{Z}_+}$. Similarly, by (5.1.1) and (5.1.3),

$$\mathbb{E}(X_i \,|\, \mathcal{F}_{i-1}) = X_{i-1} + \delta Y_{i-1} + \gamma, \quad i \in \mathbb{N},$$

and hence $(\eta_i)_{i\in\mathbb{N}}$ is a sequence of martingale differences with respect to the filtration $(\mathcal{F}_i)_{i\in\mathbb{Z}_+}$. This establishes the martingale property for $\boldsymbol{M}_{n,k}$. The next step is computing the quadratic variations. This is inevitably cumbersome, but not conceptually difficult. Applying the identities $(\boldsymbol{A}_1 \otimes \boldsymbol{A}_2)^\top = \boldsymbol{A}_1^\top \otimes \boldsymbol{A}_2^\top$ and $(\boldsymbol{A}_1 \otimes \boldsymbol{A}_2)(\boldsymbol{A}_3 \otimes \boldsymbol{A}_4) = (\boldsymbol{A}_1\boldsymbol{A}_3) \otimes (\boldsymbol{A}_2\boldsymbol{A}_4)$ (whenever the multiplications can be performed),

$$\begin{aligned}
&\mathbb{E}\left((\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})(\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})^\top \,\middle|\, \mathcal{F}_{n,k-1}\right) \\
&\qquad = \frac{1}{n}\,\mathbb{E}\left(\left(\begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix}\right)\left(\begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix}\right)^\top \,\middle|\, \mathcal{F}_{k-1}\right),
\end{aligned}$$

and, continuing,

$$
\begin{aligned}
&\mathbb{E}\left((\boldsymbol{M}_{n,k}-\boldsymbol{M}_{n,k-1})(\boldsymbol{M}_{n,k}-\boldsymbol{M}_{n,k-1})^\top \,\middle|\, \mathcal{F}_{n,k-1}\right)\\
&\qquad=\frac{1}{n}\,\mathbb{E}\left(\left(\begin{bmatrix}\varepsilon_k\\ \eta_k\end{bmatrix}\begin{bmatrix}\varepsilon_k\\ \eta_k\end{bmatrix}^\top\right)\otimes\left(\begin{bmatrix}1\\ Y_{k-1}\end{bmatrix}\begin{bmatrix}1\\ Y_{k-1}\end{bmatrix}^\top\right)\,\middle|\,\mathcal{F}_{k-1}\right)\\
&\qquad=\frac{1}{n}\,\mathbb{E}\left(\begin{bmatrix}\varepsilon_k\\ \eta_k\end{bmatrix}\begin{bmatrix}\varepsilon_k\\ \eta_k\end{bmatrix}^\top\,\middle|\,\mathcal{F}_{k-1}\right)\otimes\left(\begin{bmatrix}1\\ Y_{k-1}\end{bmatrix}\begin{bmatrix}1\\ Y_{k-1}\end{bmatrix}^\top\right),\qquad n\in\mathbb{N},\ k\in\{1,\ldots,n\}.
\end{aligned}
$$

We need to calculate the conditional expectations in the first term one by one – the reader who wishes to skip the details may find the end result in (5.3.1), (5.3.2) and (5.3.3), respectively. By (3.1.1), we have

$$
\begin{aligned}
Y_i &= \mathrm{e}^{-b}Y_{i-1}+a\int_{i-1}^{i}\mathrm{e}^{-b(i-u)}\,\mathrm{d}u+\sigma_1\int_{i-1}^{i}\mathrm{e}^{-b(i-u)}\sqrt{Y_u}\,\mathrm{d}W_u\\
&= dY_{i-1}+c+\sigma_1\int_{i-1}^{i}\mathrm{e}^{-b(i-u)}\sqrt{Y_u}\,\mathrm{d}W_u,\qquad i\in\mathbb{N},
\end{aligned}
$$

hence, by Karatzas and Shreve (1991, Proposition 3.2.10) and (3.1.4), we have

(5.3.1)

$$
\begin{aligned}
\mathbb{E}(\varepsilon_i^2\,|\,\mathcal{F}_{i-1}) &= \sigma_1^2\,\mathbb{E}\left(\left(\int_{i-1}^{i}\mathrm{e}^{-b(i-u)}\sqrt{Y_u}\,\mathrm{d}W_u\right)^2\,\middle|\,\mathcal{F}_{i-1}\right)=\sigma_1^2\int_{i-1}^{i}\mathrm{e}^{-2b(i-u)}\,\mathbb{E}(Y_u\,|\,\mathcal{F}_{i-1})\,\mathrm{d}u\\
&=\sigma_1^2\int_{i-1}^{i}\mathrm{e}^{-2b(i-u)}\mathrm{e}^{-b(u-i+1)}Y_{i-1}\,\mathrm{d}u+\sigma_1^2\int_{i-1}^{i}\mathrm{e}^{-2b(i-u)}a\int_{i-1}^{u}\mathrm{e}^{-b(u-v)}\,\mathrm{d}v\,\mathrm{d}u\\
&=\sigma_1^2Y_{i-1}\int_0^1\mathrm{e}^{-b(2-v)}\,\mathrm{d}v+\sigma_1^2a\int_0^1\int_0^u\mathrm{e}^{-b(2-v-u)}\,\mathrm{d}v\,\mathrm{d}u=:C_1Y_{i-1}+C_2.
\end{aligned}
$$

By (3.1.1) and (3.1.4), with the notation $\widetilde{W}_t:=\varrho W_t+\sqrt{1-\varrho^2}B_t$, $t\in\mathbb{R}_+$, we compute

$$
\begin{aligned}
X_i-X_{i-1} &= \int_{i-1}^{i}(\alpha-\beta Y_u)\,\mathrm{d}u+\sigma_2\int_{i-1}^{i}\sqrt{Y_u}\,\mathrm{d}\widetilde{W}_u=\alpha-\beta\int_{i-1}^{i}Y_u\,\mathrm{d}u+\sigma_2\int_{i-1}^{i}\sqrt{Y_u}\,\mathrm{d}\widetilde{W}_u\\
&=\alpha-\beta\int_{i-1}^{i}\left(\mathrm{e}^{-b(u-i+1)}Y_{i-1}+a\int_{i-1}^{u}\mathrm{e}^{-b(u-v)}\,\mathrm{d}v+\sigma_1\int_{i-1}^{u}\mathrm{e}^{-b(u-v)}\sqrt{Y_v}\,\mathrm{d}W_v\right)\mathrm{d}u\\
&\quad+\sigma_2\int_{i-1}^{i}\sqrt{Y_u}\,\mathrm{d}\widetilde{W}_u\\
&=\alpha-\beta Y_{i-1}\int_{i-1}^{i}\mathrm{e}^{-b(u-i+1)}\,\mathrm{d}u-a\beta\int_{i-1}^{i}\left(\int_{i-1}^{u}\mathrm{e}^{-b(u-v)}\,\mathrm{d}v\right)\mathrm{d}u\\
&\quad-\beta\sigma_1\int_{i-1}^{i}\left(\int_{i-1}^{u}\mathrm{e}^{-b(u-v)}\sqrt{Y_v}\,\mathrm{d}W_v\right)\mathrm{d}u+\sigma_2\int_{i-1}^{i}\sqrt{Y_u}\,\mathrm{d}\widetilde{W}_u,
\end{aligned}
$$

and, continuing,

$$\begin{aligned}
X_i - X_{i-1} &= \alpha - \beta Y_{i-1}\int_0^1 \mathrm{e}^{-bv}\,\mathrm{d}v - a\beta\int_0^1\left(\int_0^u \mathrm{e}^{-bv}\,\mathrm{d}v\right)\mathrm{d}u \\
&\quad - \beta\sigma_1\int_{i-1}^i\left(\int_{i-1}^u \mathrm{e}^{-b(u-v)}\sqrt{Y_v}\,\mathrm{d}W_v\right)\mathrm{d}u + \sigma_2\int_{i-1}^i\sqrt{Y_u}\,\mathrm{d}\widetilde{W}_u \\
&= \delta Y_{i-1} + \gamma - \beta\sigma_1\int_{i-1}^i\left(\int_{i-1}^u \mathrm{e}^{-b(u-v)}\sqrt{Y_v}\,\mathrm{d}W_v\right)\mathrm{d}u + \sigma_2\int_{i-1}^i\sqrt{Y_u}\,\mathrm{d}\widetilde{W}_u,
\end{aligned}$$

so, consequently,

$$\begin{aligned}
\mathbb{E}(\eta_i^2\,|\,\mathcal{F}_{i-1}) &= \beta^2\sigma_1^2\,\mathbb{E}\left[\left(\int_{i-1}^i\int_{i-1}^u \mathrm{e}^{-b(u-v)}\sqrt{Y_v}\,\mathrm{d}W_v\,\mathrm{d}u\right)^2\,\middle|\,\mathcal{F}_{i-1}\right] \\
&\quad + \sigma_2^2\,\mathbb{E}\left[\left(\int_{i-1}^i\sqrt{Y_u}\,\mathrm{d}\widetilde{W}_u\right)^2\,\middle|\,\mathcal{F}_{i-1}\right] \\
&\quad - 2\beta\sigma_1\sigma_2\,\mathbb{E}\left[\left(\int_{i-1}^i\int_{i-1}^u \mathrm{e}^{-b(u-v)}\sqrt{Y_v}\,\mathrm{d}W_v\,\mathrm{d}u\right)\left(\varrho\int_{i-1}^i\sqrt{Y_u}\,\mathrm{d}W_u\right)\middle|\,\mathcal{F}_{i-1}\right] \\
&\quad - 2\beta\sigma_1\sigma_2\,\mathbb{E}\left[\left(\int_{i-1}^i\int_{i-1}^u \mathrm{e}^{-b(u-v)}\sqrt{Y_v}\,\mathrm{d}W_v\mathrm{d}u\right)\left(\sqrt{1-\varrho^2}\int_{i-1}^i\sqrt{Y_u}\mathrm{d}B_u\right)\middle|\,\mathcal{F}_{i-1}\right].
\end{aligned}$$

We use Karatzas and Shreve (1991, Equation (3.2.23)) to the first, second and third terms, and Karatzas and Shreve (1991, Proposition 3.2.17) to the fourth term, together with the independence of W and B:

$$\begin{aligned}
\mathbb{E}(\eta_i^2|\mathcal{F}_{i-1}) &= \beta^2\sigma_1^2\int_{i-1}^i\int_{i-1}^i\mathbb{E}\left(\int_{i-1}^u \mathrm{e}^{-b(u-w)}\sqrt{Y_w}\,\mathrm{d}W_w\int_{i-1}^v \mathrm{e}^{-b(v-w)}\sqrt{Y_w}\,\mathrm{d}W_w\,\middle|\,\mathcal{F}_{i-1}\right)\mathrm{d}v\mathrm{d}u \\
&\quad + \sigma_2^2\int_{i-1}^i\mathbb{E}(Y_u\,|\,\mathcal{F}_{i-1})\,\mathrm{d}u \\
&\quad - 2\beta\sigma_1\sigma_2\varrho\int_{i-1}^i\mathbb{E}\left(\int_{i-1}^u \mathrm{e}^{-b(u-w)}\sqrt{Y_w}\,\mathrm{d}W_w\int_{i-1}^i\sqrt{Y_w}\,\mathrm{d}W_w\,\middle|\,\mathcal{F}_{i-1}\right)\mathrm{d}u - 0 \\
&= \beta^2\sigma_1^2\int_{i-1}^i\int_{i-1}^i\int_{i-1}^{u\wedge v}\mathrm{e}^{-b(u+v-2w)}\,\mathbb{E}(Y_w\,|\,\mathcal{F}_{i-1})\,\mathrm{d}w\,\mathrm{d}u\,\mathrm{d}v + \sigma_2^2\int_{i-1}^i\mathbb{E}(Y_u\,|\,\mathcal{F}_{i-1})\,\mathrm{d}u \\
&\quad - 2\beta\sigma_1\sigma_2\varrho\int_{i-1}^i\int_{i-1}^u \mathrm{e}^{-b(u-v)}\mathbb{E}(Y_v\,|\,\mathcal{F}_{i-1})\,\mathrm{d}v\,\mathrm{d}u.
\end{aligned}$$

Using again (3.1.4), we get

(5.3.2)

$$
\begin{aligned}
&\mathbb{E}(\eta_i^2|\mathcal{F}_{i-1}) \\
&= \beta^2\sigma_1^2 Y_{i-1}\int_{i-1}^{i}\int_{i-1}^{i}\int_{i-1}^{u\wedge v} \mathrm{e}^{-b(u+v-w-(i-1))}\,\mathrm{d}w\,\mathrm{d}v\,\mathrm{d}u \\
&\quad + a\beta^2\sigma_1^2\int_{i-1}^{i}\int_{i-1}^{i}\int_{i-1}^{u\wedge v}\int_{i-1}^{w} \mathrm{e}^{-b(u+v-w-z)}\,\mathrm{d}z\,\mathrm{d}w\,\mathrm{d}v\,\mathrm{d}u \\
&\quad + \sigma_2^2 Y_{i-1}\int_{i-1}^{i} e^{-b(u-(i-1))}\,\mathrm{d}u + a\sigma_2^2\int_{i-1}^{i}\int_{i-1}^{u}\mathrm{e}^{-b(u-v)}\,\mathrm{d}v\,\mathrm{d}u \\
&\quad - 2\beta\sigma_1\sigma_2\varrho Y_{i-1}\int_{i-1}^{i}\int_{i-1}^{u}\mathrm{e}^{-b(u-(i-1))}\,\mathrm{d}v\,\mathrm{d}u - 2a\beta\sigma_1\sigma_2\varrho\int_{i-1}^{i}\int_{i-1}^{u}\int_{i-1}^{v}\mathrm{e}^{-b(u-w)}\,\mathrm{d}w\,\mathrm{d}v\,\mathrm{d}u \\
&= \Bigg(\beta^2\sigma_1^2\int_0^1\int_0^1\int_0^{u'\wedge v'}\mathrm{e}^{-b(u'+v'-w')}\,\mathrm{d}w'\,\mathrm{d}v'\,\mathrm{d}u' \\
&\qquad - 2\beta\sigma_1\sigma_2\varrho\int_0^1\int_0^{u'}\mathrm{e}^{-bu'}\,\mathrm{d}v'\,\mathrm{d}u' + \sigma_2^2\int_0^1\mathrm{e}^{-bu'}\,\mathrm{d}u'\Bigg)Y_{i-1} \\
&\quad + a\beta^2\sigma_1^2\int_0^1\int_0^1\int_0^{u'\wedge v'}\int_0^{w'}\mathrm{e}^{-b(u'+v'-w'-z')}\,\mathrm{d}z'\,\mathrm{d}w'\,\mathrm{d}v'\,\mathrm{d}u' \\
&\quad + a\sigma_2^2\int_0^1\int_0^{u'}\mathrm{e}^{-b(u'-v')}\,\mathrm{d}v'\,\mathrm{d}u' - 2a\beta\sigma_1\sigma_2\varrho\int_0^1\int_0^{u'}\int_0^{v'}\mathrm{e}^{-b(u'-w')}\,\mathrm{d}w'\,\mathrm{d}v'\,\mathrm{d}u' \\
&=: C_3 Y_{i-1} + C_4.
\end{aligned}
$$

To calculate the off-diagonal entries in the quadratic variation, we write

$$
\begin{aligned}
&\mathbb{E}(\varepsilon_k\eta_k \mid \mathcal{F}_{k-1}) = \mathbb{E}\big((Y_k - c - dY_{k-1})(X_k - X_{k-1} - \gamma - \delta Y_{k-1}) \,\big|\, \mathcal{F}_{k-1}\big) \\
&= \mathbb{E}\Bigg(\sigma_1\int_{k-1}^{k}\mathrm{e}^{-b(k-s)}\sqrt{Y_s}\,\mathrm{d}W_s \times \\
&\qquad \times\left(-\beta\sigma_1\int_{k-1}^{k}\int_{k-1}^{u}\mathrm{e}^{-b(u-v)}\sqrt{Y_v}\,\mathrm{d}W_v\,\mathrm{d}u + \sigma_2\int_{k-1}^{k}\sqrt{Y_u}\,\mathrm{d}\widetilde{W}_u\right)\Bigg|\,\mathcal{F}_{k-1}\Bigg) \\
&= -\beta\sigma_1^2\int_{k-1}^{k}\mathbb{E}\left(\int_{k-1}^{k}\mathrm{e}^{-b(k-s)}\sqrt{Y_s}\,\mathrm{d}W_s\int_{k-1}^{u}\mathrm{e}^{-b(u-v)}\sqrt{Y_v}\,\mathrm{d}W_v\,\Bigg|\,\mathcal{F}_{k-1}\right)\mathrm{d}u \\
&\quad + \sigma_1\sigma_2\,\mathbb{E}\left(\int_{k-1}^{k}\mathrm{e}^{-b(k-s)}\sqrt{Y_s}\,\mathrm{d}W_s\int_{k-1}^{k}\sqrt{Y_u}\,\mathrm{d}\widetilde{W}_u\,\Bigg|\,\mathcal{F}_{k-1}\right).
\end{aligned}
$$

Again by Karatzas and Shreve (1991, Equation (3.2.23) and Proposition 3.2.17), we have

$$\begin{aligned}\mathbb{E}(\varepsilon_k\eta_k \mid \mathcal{F}_{k-1}) = -\beta\sigma_1^2 \int_{k-1}^{k}\int_{k-1}^{u} \mathrm{e}^{-b(k+u-2v)}\,\mathbb{E}(Y_v \mid \mathcal{F}_{k-1})\,\mathrm{d}v\,\mathrm{d}u\\ + \sigma_1\sigma_2\varrho\int_{k-1}^{k}\mathrm{e}^{-b(k-v)}\,\mathbb{E}(Y_v \mid \mathcal{F}_{k-1})\,\mathrm{d}v.\end{aligned}$$

Using (3.1.4), by an easy calculation,

(5.3.3)

$$\begin{aligned}\mathbb{E}(\varepsilon_k\eta_k \mid \mathcal{F}_{k-1}) &= -\beta\sigma_1^2 \int_{k-1}^{k}\int_{k-1}^{u} \mathrm{e}^{-b(k+u-2v)}\left(\mathrm{e}^{-b(v-k+1)}Y_{k-1} + a\int_{k-1}^{v}\mathrm{e}^{-b(v-s)}\,\mathrm{d}s\right)\mathrm{d}v\,\mathrm{d}u\\ &\quad + \sigma_1\sigma_2\varrho\int_{k-1}^{k}\mathrm{e}^{-b(k-v)}\left(\mathrm{e}^{-b(v-k+1)}Y_{k-1} + a\int_{k-1}^{v}\mathrm{e}^{-b(v-s)}\,\mathrm{d}s\right)\mathrm{d}v\\ &= \left(-\beta\sigma_1^2\int_0^1\int_0^{u'}\mathrm{e}^{-b(u'-v'+1)}\,\mathrm{d}v'\,\mathrm{d}u' + \sigma_1\sigma_2\varrho\mathrm{e}^{-b}\right)Y_{k-1}\\ &\quad - a\beta\sigma_1^2\int_0^1\int_0^{u'}\int_0^{v'}\mathrm{e}^{-b(u'-v'-s'+1)}\,\mathrm{d}s'\,\mathrm{d}v'\,\mathrm{d}u'\\ &\quad + a\sigma_1\sigma_2\varrho\int_0^1\int_0^{v'}\mathrm{e}^{-b(1-s')}\,\mathrm{d}s'\,\mathrm{d}v'\\ &=: C_5Y_{k-1} + C_6, \qquad k\in\mathbb{N}.\end{aligned}$$

Summarizing we have for the quadratic variation, by (3.2.2) and (3.2.4),

$$\begin{aligned}&\sum_{k=1}^{n}\mathbb{E}\left((\boldsymbol{M}_{n,k}-\boldsymbol{M}_{n,k-1})(\boldsymbol{M}_{n,k}-\boldsymbol{M}_{n,k-1})^\top \mid \mathcal{F}_{n,k-1}\right)\\ &\qquad = \frac{1}{n}\sum_{k=1}^{n}\begin{bmatrix}C_1Y_{k-1}+C_2 & C_5Y_{k-1}+C_6\\ C_5Y_{k-1}+C_6 & C_3Y_{k-1}+C_4\end{bmatrix}\otimes\begin{bmatrix}1 & Y_{k-1}\\ Y_{k-1} & Y_{k-1}^2\end{bmatrix}\\ &\qquad = \frac{1}{n}\sum_{k=1}^{n}\begin{bmatrix}C_1 & C_5\\ C_5 & C_3\end{bmatrix}\otimes\begin{bmatrix}Y_{k-1} & Y_{k-1}^2\\ Y_{k-1}^2 & Y_{k-1}^3\end{bmatrix} + \frac{1}{n}\sum_{k=1}^{n}\begin{bmatrix}C_2 & C_6\\ C_6 & C_4\end{bmatrix}\otimes\begin{bmatrix}1 & Y_{k-1}\\ Y_{k-1} & Y_{k-1}^2\end{bmatrix}\\ &\qquad \xrightarrow{\text{a.s.}} \begin{bmatrix}C_1 & C_5\\ C_5 & C_3\end{bmatrix}\otimes\begin{bmatrix}\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2)\\ \mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3)\end{bmatrix} + \begin{bmatrix}C_2 & C_6\\ C_6 & C_4\end{bmatrix}\otimes\begin{bmatrix}1 & \mathbb{E}(Y_\infty)\\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2)\end{bmatrix} =: \boldsymbol{D},\end{aligned} \tag{5.3.4}$$

where the 4×4 limit matrix $\boldsymbol{D}$ is necessarily symmetric and positive semi-definite (indeed, the limit of positive semi-definite matrices is positive semi-definite). For the definitions of $C_i, i = 1,\ldots,6$, see (5.3.1), (5.3.2) and (5.3.3).

We complete the proof by checking the Lindeberg condition (3.3.3). Since

$$\|\boldsymbol{x}\|^2 \mathbb{1}_{\{\|\boldsymbol{x}\| \geqslant \varepsilon\}} \leqslant \frac{\|\boldsymbol{x}\|^4}{\varepsilon^2} \mathbb{1}_{\{\|\boldsymbol{x}\| \geqslant \varepsilon\}} \leqslant \frac{\|\boldsymbol{x}\|^4}{\varepsilon^2}, \qquad \boldsymbol{x} \in \mathbb{R}^4, \quad \varepsilon \in \mathbb{R}_{++},$$

and $\|\boldsymbol{x}\|^4 = (x_1^2 + x_2^2 + x_3^2 + x_4^2)^2 \leqslant 4(x_1^4 + x_2^4 + x_3^4 + x_4^4)$, $x_1, x_2, x_3, x_4 \in \mathbb{R}$, it is enough to check that

$$\frac{1}{n^2} \sum_{k=1}^{n} \big(\mathbb{E}(\varepsilon_k^4 \,|\, \mathcal{F}_{k-1}) + Y_{k-1}^4 \, \mathbb{E}(\varepsilon_k^4 \,|\, \mathcal{F}_{k-1}) + \mathbb{E}(\eta_k^4 \,|\, \mathcal{F}_{k-1}) + Y_{k-1}^4 \, \mathbb{E}(\eta_k^4 \,|\, \mathcal{F}_{k-1}) \big)$$
$$= \frac{1}{n^2} \sum_{k=1}^{n} \mathbb{E}((1 + Y_{k-1}^4)(\varepsilon_k^4 + \eta_k^4) \,|\, \mathcal{F}_{k-1}) \xrightarrow{\mathbb{P}} 0 \qquad \text{as } n \to \infty.$$

Instead of convergence in probability, we show convergence in L^1, i.e., we check that

$$\frac{1}{n^2} \sum_{k=1}^{n} \mathbb{E}((1 + Y_{k-1}^4)(\varepsilon_k^4 + \eta_k^4)) \to 0 \qquad \text{as } n \to \infty.$$

Clearly, it is enough to show that

$$\sup_{k \in \mathbb{N}} \mathbb{E}((1 + Y_{k-1}^4)(\varepsilon_k^4 + \eta_k^4)) < \infty.$$

By the Cauchy–Schwarz inequality,

$$\mathbb{E}((1 + Y_{k-1}^4)(\varepsilon_k^4 + \eta_k^4)) \leqslant \sqrt{\mathbb{E}((1 + Y_{k-1}^4)^2) \, \mathbb{E}((\varepsilon_k^4 + \eta_k^4)^2)} \leqslant \sqrt{2} \sqrt{\mathbb{E}((1 + Y_{k-1}^4)^2) \, \mathbb{E}(\varepsilon_k^8 + \eta_k^8)}$$

for all $k \in \mathbb{N}$. Since, as stated in (3.1.3),

$$\sup_{t \in \mathbb{R}_+} \mathbb{E}(Y_t^\kappa) < \infty, \qquad \kappa \in \mathbb{R}_+, \tag{5.3.5}$$

it remains to check that $\sup_{k \in \mathbb{N}} \mathbb{E}(\varepsilon_k^8 + \eta_k^8) < \infty$. Since, by the power mean inequality,

$$\mathbb{E}(\varepsilon_k^8) = \mathbb{E}(|Y_k - dY_{k-1} - c|^8) \leqslant \mathbb{E}((Y_k + dY_{k-1} + c)^8) \leqslant 3^7 \, \mathbb{E}(Y_k^8 + d^8 Y_{k-1}^8 + c^8), \qquad k \in \mathbb{N},$$

using (5.3.5), we have $\sup_{k \in \mathbb{N}} \mathbb{E}(\varepsilon_k^8) < \infty$. Using (3.1.1),

$$\mathbb{E}(\eta_k^8) = \mathbb{E}((X_k - X_{k-1} - \gamma - \delta Y_{k-1})^8)$$
$$= \mathbb{E}\left(\left(\alpha - \beta \int_{k-1}^{k} Y_u \, \mathrm{d}u + \sigma_2 \varrho \int_{k-1}^{k} \sqrt{Y_u} \, \mathrm{d}W_u + \sigma_2 \sqrt{1 - \varrho^2} \int_{k-1}^{k} \sqrt{Y_u} \, \mathrm{d}B_u - \gamma - \delta Y_{k-1} \right)^8 \right),$$

and using the power mean inequality again,

$$\mathbb{E}(\eta_k^8) \leqslant 6^7\,\mathbb{E}\Bigg(\alpha^8 + \beta^8\left(\int_{k-1}^k Y_u\,\mathrm{d}u\right)^8 + \sigma_2^8\varrho^8\left(\int_{k-1}^k \sqrt{Y_u}\,\mathrm{d}W_u\right)^8$$
$$+ \sigma_2^8(1-\varrho^2)^4\left(\int_{k-1}^k \sqrt{Y_u}\,\mathrm{d}B_u\right)^8 + \delta^8 Y_{k-1}^8 + \gamma^8\Bigg).$$

By Jensen's inequality and (5.3.5),

$$\begin{aligned}\sup_{k\in\mathbb{N}}\mathbb{E}\left(\left(\int_{k-1}^k Y_u\,\mathrm{d}u\right)^8\right) &\leqslant \sup_{k\in\mathbb{N}}\mathbb{E}\left(\int_{k-1}^k Y_u^8\,\mathrm{d}u\right) = \sup_{k\in\mathbb{N}}\int_{k-1}^k \mathbb{E}(Y_u^8)\,\mathrm{d}u\\ &\leqslant \left(\sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t^8)\right)\left(\sup_{k\in\mathbb{N}}\int_{k-1}^k 1\,\mathrm{d}u\right) = \sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t^8) < \infty.\end{aligned} \tag{5.3.6}$$

By the SDE (3.0.1), the power mean inequality, and (5.3.6),

$$\begin{aligned}\mathbb{E}\left(\left(\int_{k-1}^k \sqrt{Y_u}\,\mathrm{d}W_u\right)^8\right) &\leqslant \frac{1}{\sigma_1^8}\mathbb{E}\left(\left(Y_k - Y_{k-1} - a - b\int_{k-1}^k Y_u\,\mathrm{d}u\right)^8\right)\\ &\leqslant \frac{4^7}{\sigma_1^8}\mathbb{E}\left(Y_k^8 + Y_{k-1}^8 + a^8 + b^8\left(\int_{k-1}^k Y_u\,\mathrm{d}u\right)^8\right)\\ &\leqslant \frac{4^7}{\sigma_1^8}\left(2\sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t^8) + a^8 + b^8\sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t^8)\right) < \infty.\end{aligned}$$

Further, using that the conditional distribution of $\int_{k-1}^k \sqrt{Y_u}\,\mathrm{d}B_u$ given $(Y_u)_{u\in[0,k]}$ is normal with mean 0 and variance $\int_{k-1}^k Y_u\,\mathrm{d}u$, we have

$$\mathbb{E}\left(\left(\int_{k-1}^k \sqrt{Y_u}\,\mathrm{d}B_u\right)^8 \,\Bigg|\, (Y_u)_{u\in[0,k]}\right) = 105\left(\int_{k-1}^k Y_u\,\mathrm{d}u\right)^4, \qquad k\in\mathbb{N},$$

and consequently

$$\mathbb{E}\left(\left(\int_{k-1}^k \sqrt{Y_u}\,\mathrm{d}B_u\right)^8\right) = 105\,\mathbb{E}\left(\left(\int_{k-1}^k Y_u\,\mathrm{d}u\right)^4\right), \qquad k\in\mathbb{N}.$$

Hence, similarly to (5.3.6), we have

$$\sup_{k\in\mathbb{N}}\mathbb{E}\left(\left(\int_{k-1}^k \sqrt{Y_u}\,\mathrm{d}B_u\right)^8\right) \leqslant 105\sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t^4) < \infty,$$

which yields that $\sup_{k\in\mathbb{N}}\mathbb{E}(\eta_k^8) < \infty$. □

5.3.2 Lemma. *Under the conditions of Theorem 5.2.1 and $\boldsymbol{M}_{n,k}$ defined in* (5.2.4)*, we have*

$$n^{-1/2}\boldsymbol{M}_{n,n} \xrightarrow{\text{a.s.}} \mathbf{0} \quad \textit{componentwise.}$$

Proof. First let us observe that

$$\boldsymbol{M}_n := n^{1/2}\boldsymbol{M}_{n,n} = \sum_{i=1}^{n} \begin{bmatrix} \varepsilon_i \\ \eta_i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}$$

is a martingale. The reasoning is the same as the beginning of the proof of Lemma 5.3.2. We will use Theorem 3.3.3 componentwise to show that

$$\frac{\boldsymbol{M}_n}{n} \xrightarrow{\text{a.s.}} \mathbf{0}$$

componentwise, which is our statement. Let us recall (5.3.4) and introduce

$$\boldsymbol{D}_n := \sum_{k=1}^{n} \mathbb{E}\left((\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})(\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})^\top \,|\, \mathcal{F}_{n,k-1}\right).$$

Now, for the components of $\boldsymbol{M}_n$, we have $\langle M^{(i)}\rangle_n = D_n^{i,i}$, that is, the i-th element on the main diagonal in D_n, for $i = 1, 2, 3, 4$. Hence,

$$\frac{M_n^{(i)}}{n} = \frac{M_n^{(i)}}{D_n} \cdot \frac{D_n}{n} \xrightarrow{\text{a.s.}} 0 \cdot D^{(i,i)} = 0,$$

by using the convergence from (5.3.4), and $D^{(i,i)}$ denoting the i-th element in the main diagonal of D. □

5.3.3 Lemma. *Under the conditions of Theorem 5.2.1, $\boldsymbol{G}$ is positive definite.*

Proof. Expanding the definition of $\boldsymbol{G}$,

$$\begin{aligned}
\boldsymbol{G} = &\left(\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \otimes \left(\begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \\ \mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix}\right)\right) \times \\
&\times \left(\boldsymbol{E}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}\right) \\
&+ \left(\begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \left(\begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}\right)\right) \times \\
&\times \left(\boldsymbol{E}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}\right),
\end{aligned}$$

from which

$$\boldsymbol{G} = \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \otimes \left(\begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \\ \mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \right)$$
$$+ \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \left(\begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \right).$$

Writing out the inverses,

$$\begin{aligned} \boldsymbol{G} &= \frac{1}{(\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)^2} \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \\ &\quad \otimes \left(\begin{bmatrix} \mathbb{E}(Y_\infty^2) & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & 1 \end{bmatrix} \begin{bmatrix} \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \\ \mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix} \begin{bmatrix} \mathbb{E}(Y_\infty^2) & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & 1 \end{bmatrix} \right) \\ &\quad + \frac{1}{\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2} \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \begin{bmatrix} \mathbb{E}(Y_\infty^2) & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & 1 \end{bmatrix} \\ &= \frac{1}{(\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)^2} \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \\ &\quad \otimes \begin{bmatrix} -\mathbb{E}(Y_\infty)((\mathbb{E}(Y_\infty^2))^2 - \mathbb{E}(Y_\infty)\,\mathbb{E}(Y_\infty^3)) & (\mathbb{E}(Y_\infty^2))^2 - \mathbb{E}(Y_\infty)\,\mathbb{E}(Y_\infty^3) \\ (\mathbb{E}(Y_\infty^2))^2 - \mathbb{E}(Y_\infty)\,\mathbb{E}(Y_\infty^3) & \mathbb{E}(Y_\infty^3) - 2\,\mathbb{E}(Y_\infty)\,\mathbb{E}(Y_\infty^2) + (\mathbb{E}(Y_\infty))^3 \end{bmatrix} \\ &\quad + \frac{1}{\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2} \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \begin{bmatrix} \mathbb{E}(Y_\infty^2) & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & 1 \end{bmatrix}. \end{aligned}$$

All in all,

$$\boldsymbol{G} = \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \otimes \begin{bmatrix} \frac{a(2a+\sigma_1^2)}{b\sigma_1^2} & -\frac{2a+\sigma_1^2}{\sigma_1^2} \\ -\frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{a\sigma_1^2} \end{bmatrix} + \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \begin{bmatrix} \frac{2a+\sigma_1^2}{\sigma_1^2} & -\frac{2b}{\sigma_1^2} \\ -\frac{2b}{\sigma_1^2} & \frac{2b^2}{a\sigma_1^2} \end{bmatrix}. \tag{5.3.7}$$

Indeed, by (3.2.2), an easy calculation shows that

$$\begin{aligned} \big(\mathbb{E}(Y_\infty)\,\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2\big)\,\mathbb{E}(Y_\infty) &= \frac{a^3\sigma_1^2}{4b^5}(2a+\sigma_1^2), \\ \mathbb{E}(Y_\infty)\,\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2 &= \frac{a^2\sigma_1^2}{4b^4}(2a+\sigma_1^2), \\ \mathbb{E}(Y_\infty^3) - 2\,\mathbb{E}(Y_\infty)\,\mathbb{E}(Y_\infty^2) + (\mathbb{E}(Y_\infty))^3 &= \frac{a\sigma_1^2}{2b^3}(a+\sigma_1^2), \\ \mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2 &= \frac{a\sigma_1^2}{2b^2}. \end{aligned}$$

To show the statement from here, it is enough to check that

(i) the matrix
$$\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix}$$
is positive definite,

(ii) the matrices
$$\begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix}, \qquad \begin{bmatrix} \frac{a(2a+\sigma_1^2)}{b\sigma_1^2} & -\frac{2a+\sigma_1^2}{\sigma_1^2} \\ -\frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{a\sigma_1^2} \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} \frac{2a+\sigma_1^2}{\sigma_1^2} & -\frac{2b}{\sigma_1^2} \\ -\frac{2b}{\sigma_1^2} & \frac{2b^2}{a\sigma_1^2} \end{bmatrix}$$
are positive semi-definite.

Indeed, the sum of a positive definite and a positive semi-definite square matrix is positive definite, the Kronecker product of positive semi-definite matrices is positive semi-definite and the Kronecker product of positive definite matrices is positive definite (as a consequence of the fact that the eigenvalues of the Kronecker product of two square matrices are the product of the eigenvalues of the two square matrices in question including multiplicities). The positive semi-definiteness of the matrices
$$\begin{bmatrix} \frac{a(2a+\sigma_1^2)}{b\sigma_1^2} & -\frac{2a+\sigma_1^2}{\sigma_1^2} \\ -\frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{a\sigma_1^2} \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} \frac{2a+\sigma_1^2}{\sigma_1^2} & -\frac{2b}{\sigma_1^2} \\ -\frac{2b}{\sigma_1^2} & \frac{2b^2}{a\sigma_1^2} \end{bmatrix}$$
readily follows, since $\frac{a(2a+\sigma_1^2)}{b\sigma_1^2} > 0$, $\frac{2a+\sigma_1^2}{\sigma_1^2} > 0$, and the determinant of the matrices in question are $\frac{2a+\sigma_1^2}{\sigma_1^2} > 0$ and $\frac{2b}{a\sigma_1^2} > 0$, respectively. Next, we prove that the matrices
$$\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} C_2 & C_4 \\ C_4 & C_6 \end{bmatrix}$$
are positive semi-definite. Since $\mathbb{P}(Y_0 = y_0) = 1$, we have $\mathbb{E}(\varepsilon_1^2 \,|\, \mathcal{F}_0) = C_1 y_0 + C_2$, $\mathbb{E}(\eta_1^2 \,|\, \mathcal{F}_0) = C_3 y_0 + C_4$, and $\mathbb{E}(\varepsilon_1 \eta_1 \,|\, \mathcal{F}_0) = C_5 y_0 + C_6$ $\mathbb{P}$-almost surely, hence
$$\mathbb{E}(\varepsilon_1^2)\,\mathbb{E}(\eta_1^2) - \big(\mathbb{E}(\varepsilon_1\eta_1)\big)^2 = (C_1C_3 - C_5^2)y_0^2 + (C_1C_4 + C_2C_3 - 2C_5C_6)y_0 + C_2C_4 - C_6^2.$$
By the Cauchy–Schwarz inequality,
$$\mathbb{E}(\varepsilon_1^2)\,\mathbb{E}(\eta_1^2) - \big(\mathbb{E}(\varepsilon_1\eta_1)\big)^2 \geqslant 0,$$
hence, by setting an arbitrary initial value $Y_0 = y_0 \in \mathbb{R}_+$, we obtain $C_1C_3 - C_5^2 \geqslant 0$ and $C_2C_4 - C_6^2 \geqslant 0$.

Thus, both matrices

$$\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} C_2 & C_4 \\ C_4 & C_6 \end{bmatrix}$$

are positive semi-definite, since $C_1 > 0$ and $C_2 > 0$. Now we turn to check that

$$\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix}$$

is positive definite. Since $C_1 > 0$, this is equivalent to showing that $C_1C_3 - C_5^2 > 0$. Recalling the definition of the constants from (5.3.1), (5.3.2) and (5.3.3), we have

$$\begin{aligned}
C_1 &= \sigma_1^2 \int_0^1 \mathrm{e}^{-b(2-v)}\,\mathrm{d}v = \sigma_1^2 \mathrm{e}^{-2b}\frac{\mathrm{e}^b - 1}{b},\\
C_3 &= \beta^2\sigma_1^2 \int_0^1\int_0^1\int_0^{u'\wedge v'} \mathrm{e}^{-b(u'+v'-w')}\,\mathrm{d}w'\,\mathrm{d}v'\,\mathrm{d}u'\\
&\quad - 2\beta\sigma_1\sigma_2\varrho \int_0^1\int_0^{u'} \mathrm{e}^{-bu'}\,\mathrm{d}v'\,\mathrm{d}u' + \sigma_2^2\int_0^1 \mathrm{e}^{-bu'}\,\mathrm{d}u'\\
&= b^{-3}\left(2\mathrm{e}^{-b}\beta^2\sigma_1^2(\sinh b - b) + 2b\beta\varrho\sigma_1\sigma_2((1+b)\mathrm{e}^{-b} - 1) + b^2\sigma_2^2(1-\mathrm{e}^{-b})\right),\\
C_5 &= -\beta\sigma_1^2\int_0^1\int_0^{u'} \mathrm{e}^{-b(u'-v'+1)}\,\mathrm{d}v'\,\mathrm{d}u' + \sigma_1\sigma_2\varrho\mathrm{e}^{-b}\\
&= b^{-2}\sigma_1\mathrm{e}^{-b}\left(-\mathrm{e}^{-b}\beta\sigma_1(1+(b-1)\mathrm{e}^b) + \varrho\sigma_2 b^2\right),
\end{aligned}$$

thus we have

$$\begin{aligned}
C_1C_3 - C_5^2 = b^{-4}\mathrm{e}^{-2b}\sigma_1^2\Big(&2b(2+b^2)\beta\varrho\sigma_1\sigma_2 + 2(\beta^2\sigma_1^2 - 2b\beta\varrho\sigma_1\sigma_2 + b^2\sigma_2^2)\cosh b\\
&- (2+b^2)\beta^2\sigma_1^2 - b^2(2+b^2\varrho^2)\sigma_2^2\Big).
\end{aligned}$$

Consequently, using that

$$\cosh b = \sum_{k=0}^{\infty}\frac{b^{2k}}{(2k)!} > 1 + \frac{b^2}{2}$$

and that

$$\beta^2\sigma_1^2 - 2b\beta\varrho\sigma_1\sigma_2 + b^2\sigma_2^2 = (\beta\sigma_1 - b\varrho\sigma_2)^2 + b^2(1-\varrho^2)\sigma_2^2 > 0,$$

we have

$$
\begin{aligned}
C_1C_3 - C_5^2 > b^{-4}\mathrm{e}^{-2b}\sigma_1^2\Big(&4b\beta\varrho\sigma_1\sigma_2 + 2b^3\beta\varrho\sigma_1\sigma_2 + 2\beta^2\sigma_1^2 + b^2\beta^2\sigma_1^2 - 4b\beta\varrho\sigma_1\sigma_2 - 2b^3\beta\varrho\sigma_1\sigma_2\\
&+ 2b^2\sigma_2^2 + b^4\sigma_2^2 - 2\beta^2\sigma_1^2 - b^2\beta^2\sigma_1^2 - 2b^2\sigma_2^2 - b^4\varrho^2\sigma_2^2\Big)\\
= b^{-4}\mathrm{e}^{-2b}\sigma_1^2(b^4(1-\varrho^2)\sigma_2^2) > 0.&
\end{aligned}
$$

This concludes the proof. □

5.4 Asymptotic results for the untransformed parameters

So far we have obtained the limit distribution of the CLSE's of the transformed parameters (c, d, γ, δ). A natural estimate of (a, b, α, β) can be obtained from (5.1.2) using relation (5.1.3) detailed as follows. Calculating the integrals in (5.1.3) in the subcritical case, let us introduce the function $g : \mathbb{R}_{++}^2 \times \mathbb{R}^2 \to \mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2$,

$$
(5.4.1) \qquad g(a, b, \alpha, \beta) := \begin{bmatrix} ab^{-1}(1-\mathrm{e}^{-b}) \\ \mathrm{e}^{-b} \\ \alpha - a\beta b^{-2}(\mathrm{e}^{-b} - 1 + b) \\ -\beta b^{-1}(1-\mathrm{e}^{-b}) \end{bmatrix} = \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix}, \qquad (a, b, \alpha, \beta) \in \mathbb{R}_{++}^2 \times \mathbb{R}^2.
$$

Note that g is bijective, with the inverse

$$
(5.4.2) \qquad g^{-1}(c, d, \gamma, \delta) = \begin{bmatrix} -c\frac{\log d}{1-d} \\ -\log d \\ \gamma - c\delta\frac{d-1-\log d}{(1-d)^2} \\ \delta\frac{\log d}{1-d} \end{bmatrix} = \begin{bmatrix} a \\ b \\ \alpha \\ \beta \end{bmatrix}, \qquad (c, d, \gamma, \delta) \in \mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2.
$$

Indeed, for all $(c, d, \gamma, \delta) \in \mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2$, we have

$$
\begin{aligned}
\alpha &= \gamma + a\beta b^{-2}(\mathrm{e}^{-b} - 1 + b) = \gamma + (-c)\frac{\log d}{1-d}\delta\frac{\log d}{1-d}(-\log d)^{-2}(d - 1 - \log d)\\
&= \gamma - c\delta\frac{d-1-\log d}{(1-d)^2}.
\end{aligned}
$$

Under the conditions of Theorem 5.2.1 the CLSE $(\widehat{c}_n, \widehat{d}_n, \widehat{\gamma}_n, \widehat{\delta}_n)$ of (c, d, γ, δ) is strongly consistent, hence in the subcritical case $(\widehat{c}_n, \widehat{d}_n, \widehat{\gamma}_n, \widehat{\delta}_n)$ fall into the set $\mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2$ for sufficiently large $n \in \mathbb{N}$ with probability one. Hence, in the subcritical case, one can introduce a natural estimator of (a, b, α, β) based on discrete time observations by applying the inverse of g to the CLSE of (c, d, γ, δ), i.e.,

$$(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n) := g^{-1}(\widehat{c}_n, \widehat{d}_n, \widehat{\gamma}_n, \widehat{\delta}_n) \tag{5.4.3}$$

for sufficiently large $n \in \mathbb{N}$ with probability one.

5.4.1 Remark. We would like to stress the point that the estimator of (a, b, α, β) introduced in (5.4.3) exists only for sufficiently large $n \in \mathbb{N}$ with probability of 1. However, as all our results are asymptotic, this will not cause a problem. From the considerations before this remark, we obtain

(5.4.4)
$$\left(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n\right) = \operatorname*{arg\,min}_{(a,b,\alpha,\beta)\in\mathbb{R}^2_{++}\times\mathbb{R}^2} \sum_{i=1}^{n} \left[(Y_i - dY_{i-1} - c)^2 + (X_i - X_{i-1} - \gamma - \delta Y_{i-1})^2\right]$$

for sufficiently large $n \in \mathbb{N}$ with probability one. We note that $\left(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n\right)$ does not necessarily provides a CLSE of (a, b, α, β), since in (5.4.4) one takes the infimum only on the set $\mathbb{R}^2_{++} \times \mathbb{R}^2$ instead of $\mathbb{R}^4$. However, this is a relatively minor point. □

5.4.2 Theorem. *Under the conditions of Theorem 5.2.1 the sequence* $\left(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n\right)$, $n \in \mathbb{N}$, *is strongly consistent and asymptotically normal, i.e.,*

$$(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n) \xrightarrow{\text{a.s.}} (a, b, \alpha, \beta) \qquad \text{as } n \to \infty,$$

and

$$\sqrt{n} \begin{bmatrix} \widehat{a}_n - a \\ \widehat{b}_n - b \\ \widehat{\alpha}_n - \alpha \\ \widehat{\beta}_n - \beta \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_4\left(0, \boldsymbol{J}\boldsymbol{G}\boldsymbol{J}^\top\right) \qquad \text{as } n \to \infty,$$

where $\boldsymbol{G} \in \mathbb{R}^{2\times 2}$ *is a symmetric, positive definite matrix given in* (5.3.7) *and*

$$\boldsymbol{J} := \begin{bmatrix} -\frac{\log d}{1-d} & -c\frac{\log d - 1 + d^{-1}}{(1-d)^2} & 0 & 0 \\ 0 & -\frac{1}{d} & 0 & 0 \\ \delta\frac{\log d + 1 - d}{(1-d)^2} & c\delta\frac{2\log d - d + d^{-1}}{(1-d)^3} & 1 & c\frac{\log d + 1 - d}{(1-d)^2} \\ 0 & \delta\frac{\log d - 1 + d^{-1}}{(1-d)^2} & 0 & \frac{\log d}{1-d} \end{bmatrix}$$

with c, d and δ *given in* (5.1.3).

Proof. The strong consistency of $(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n)$, $n \in \mathbb{N}$, follows from the strong consistency of the CLSE of (c, d, γ, δ) proved in Theorem 5.2.1 and the continuity of g^{-1}.

For the second part of the theorem we use Theorem 5.2.1, and the so-called delta method (see, e.g., Lehmann and Romano, 2009, Theorem 11.2.14). Indeed, one can extend the function g^{-1} to be defined on $\mathbb{R}^4$ instead of $\mathbb{R}_{++} \times (0,1) \times \mathbb{R}^2$ (e.g., let it be zero on the complement of $\mathbb{R}_{++} \times (0,1) \times \mathbb{R}^2$). Even with this extension, $(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n)$ takes the form given in (5.4.3), and the Jacobian of g^{-1} at $(c, d, \gamma, \delta) \in \mathbb{R}_{++} \times (0,1) \times \mathbb{R}^2$ is clearly $\boldsymbol{J}$. $\square$

Appendix A

Results of theoretical interest for INAR(p)

A.1 Invertibility of the matrices $\boldsymbol{Q}_n$, $\boldsymbol{Q}'$ and $\boldsymbol{Q}''$

In (2.4.1) we assumed that the matrix $\boldsymbol{Q}_n$ is invertible, and similarly, in designing the one-sided tests we assumed that $\boldsymbol{Q}'$ and $\boldsymbol{Q}''$ are positive definite. The following two lemmas will show that these assumptions are correct.

A.1.1 Lemma. *For a homogeneous INAR(p) process with $\mu > 0$, for which either $\alpha_q \in (0, 1)$ for some $q \in \{1, 2, \ldots, p\}$ or $\sigma > 0$, we have*

$$\mathbb{P}(\boldsymbol{Q}_n \text{ is singular}) \to 0.$$

Proof. Since

$$\boldsymbol{Q}_n = \sum_{i=1}^{n} \begin{bmatrix} \boldsymbol{X}_{i-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{i-1} \\ 1 \end{bmatrix}^\top$$

is a sum of positive semidefinite matrices, it is positive semidefinite itself. Therefore, its singularity is equivalent to the condition that for some $\boldsymbol{0} \neq \boldsymbol{v} \in \mathbb{R}^{p+1}$ and every index $i \in \{1, \ldots, n\}$, we have

$$\boldsymbol{v}^\top \begin{bmatrix} \boldsymbol{X}_{i-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{i-1} \\ 1 \end{bmatrix}^\top \boldsymbol{v} = \boldsymbol{0},$$

which is equivalent to the condition that the linear span of

$$\left\{ \begin{bmatrix} \boldsymbol{X}_{i-1} \\ 1 \end{bmatrix}, i = 1, \ldots, n \right\}$$

is a proper subspace of $\mathbb{R}^{p+1}$. Now, using the continuity of probability, our statement is equivalent to the following:

$$\mathbb{P}\left(\operatorname{span}\left\{\begin{bmatrix}\boldsymbol{X}_{i-1}\\ 1\end{bmatrix}, i\in\mathbb{N}\right\} < \mathbb{R}^{p+1}\right) = 0, \tag{A.1.1}$$

where $<$ denotes proper subspace. For simplicity, throughout the proof we will use the notation

$$\boldsymbol{Y}_i = \begin{bmatrix}\boldsymbol{X}_{i-1}\\ 1\end{bmatrix},\ i\in\mathbb{N}.$$

It is clear that all values of $(\boldsymbol{Y}_i)_{i\in\mathbb{N}}$ fall into $\mathbb{Z}_+^p\times\{1\}$. We introduce the following notation for the set of spaces that can be spanned by the values of the process:

$$\mathcal{S} := \{S < \mathbb{R}^{p+1} : S = \operatorname{span}\{\boldsymbol{y}_1,\boldsymbol{y}_2,\dots,\boldsymbol{y}_n\},\ \boldsymbol{y}_1,\boldsymbol{y}_2,\dots,\boldsymbol{y}_n\in\mathbb{Z}_+^p\times\{1\},\ n\in\mathbb{N}\}. \tag{A.1.2}$$

We can notice that $\mathcal{S}$ is countable. Indeed, every generating system of a subspace contains a basis, therefore every subspace $S\in\mathcal{S}$ has a basis whose elements are from $\mathbb{Z}_+^p\times\{1\}$. Such a basis is from $(\mathbb{Z}_+^p\times\{1\})^k$ where $k=\dim S$, and $0\leqslant k\leqslant p$, and of course a basis corresponds to only one subspace. Now, since $\mathbb{Z}_+^p\times\{1\}$ is countable, $(\mathbb{Z}_+^p\times\{1\})^k$ is also countable for any $k\in\mathbb{N}$, therefore $\cup_{k=0}^p(\mathbb{Z}_+^p\times\{1\})^k$ is also countable, and so is $\mathcal{S}$.

Now we reformulate the event in (A.1.1):

(A.1.3)
$$\begin{aligned}\left\{\operatorname{span}\{\boldsymbol{Y}_i, i\in\mathbb{N}\} < \mathbb{R}^{p+1}\right\} &= \bigcup_{S<\mathbb{R}^{p+1}}\{\operatorname{span}\{\boldsymbol{Y}_i, i\in\mathbb{N}\} = S\} = \bigcup_{S\in\mathcal{S}}\{\operatorname{span}\{\boldsymbol{Y}_i, i\in\mathbb{N}\} = S\}\\ &\subseteq \bigcup_{S\in\mathcal{S}}\{\operatorname{span}\{\boldsymbol{Y}_i, i\in\mathbb{N}\}\subseteq S\}.\end{aligned}$$

Since the last union is countable, we can apply σ-subadditivity to show (A.1.1) if we can prove

$$\mathbb{P}\left(\operatorname{span}\{\boldsymbol{Y}_i, i\in\mathbb{N}\}\subseteq S\right) = \lim_{n\to\infty}\mathbb{P}\left(\operatorname{span}\{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\dots,\boldsymbol{Y}_n\}\subseteq S\right) = 0,\quad \forall S\in\mathcal{S}. \tag{A.1.4}$$

Here the first equality is trivial by the continuity of probability; it is the second equality which requires a more detailed proof. The first step in the proof of (A.1.4) relies on the mechanism by which the components of $\boldsymbol{Y}_{i+1}$ can be obtained from those of $\boldsymbol{Y}_i$. For a fixed $S\in\mathcal{S}$ the elements of S can be viewed as the solutions of a homogeneous system of independent linear equations, i.e., $\boldsymbol{y}\in S$ if and only if

$$\sum_{j=1}^{p+1}\lambda_{i,j}y^{(j)} = 0,\quad i=1,2,\dots,p+1-\dim S. \tag{A.1.5}$$

This representation is not unique, but we can fix one such representation. Now let us introduce

$$K(S) := \min\{j \in \{1, 2, \ldots, p\} : \max_i |\lambda_{i,j}| > 0\} \text{ and}$$

$$c(s) := \min\{i \in \{1, 2, \ldots, p + 1 - \dim S\} : |\lambda_{i,K(S)}| > 0\}.$$

This notation means that $K(S)$ is the first column index for which a nonzero coefficient appears in some equation in (A.1.5) and the first nonzero coefficient in the $c(s)$-th equation has index $K(S)$. We also note that $K(S) = p + 1$ is impossible because that would mean that the only equation is $y^{(p+1)} = 0$, which does not hold for any element of $\mathbb{Z}_+^p \times \{1\}$.

Let us now fix an arbitrary $i \in \mathbb{N}$ and $\omega \in \Omega$ from our underlying probability space such that $\boldsymbol{Y}_i(\omega) = \boldsymbol{y} = (y^{(1)}, y^{(2)}, \ldots, y^{(p)}, 1)^\top$. Then we have

$$\boldsymbol{Y}_{i+K(S)}(\omega) = \Big(X_{i+K(S)-1}(\omega), \ldots, X_i(\omega), y^{(1)}, \ldots, y^{(p-K(S))}, 1\Big)^\top \quad \text{(see (2.2.2))}.$$

Hence, for $\boldsymbol{Y}_{i+K(S)}(\omega) \in S$ to hold, it is necessary (but usually not sufficient) that $\boldsymbol{Y}_{i+K(S)}(\omega)$ satisfy the $c(s)$-th equation in (A.1.5), i.e.,

$$\sum_{j=1}^{K(S)} \lambda_{c(s),j} X_{i+K(S)-j}(\omega) + \sum_{j=K(S)+1}^{p} \lambda_{c(s),j} y^{(j-K(S))} + \lambda_{c(s),p+1} = 0$$

$$\Leftrightarrow \lambda_{c(s),K(S)} X_i(\omega) + \sum_{j=K(S)+1}^{p} \lambda_{c(s),j} y^{(j-K(S))} + \lambda_{c(s),p+1} = 0.$$

This linear equation has a unique solution for $X_i(\omega)$ because $\lambda_{c(s),K(S)} \neq 0$. Let us denote this unique solution by $m(\boldsymbol{y}, S)$ (by simple algebraic considerations one can see that this quantity does not depend on the representation in (A.1.5), but this is not necessary to our proof). Therefore, if $X_i(\omega) \neq m(\boldsymbol{y}, S)$, then $\omega \notin \{\boldsymbol{Y}_{i+K(S)} \in S, \boldsymbol{Y}_i = \boldsymbol{y}\}$, hence

$$\{\boldsymbol{Y}_{i+K(S)} \in S, \boldsymbol{Y}_i = \boldsymbol{y}\} \subseteq \{X_i = m(\boldsymbol{y}, S), \boldsymbol{Y}_i = \boldsymbol{y}\} \quad \forall i \in \mathbb{N}, \quad \forall \boldsymbol{y} \in \mathbb{Z}_+^p \times \{1\}.$$

If $m(\boldsymbol{y}, S) \notin \mathbb{N}$, then we have $\{\boldsymbol{Y}_{i+K(S)} \in S, \boldsymbol{Y}_i = \boldsymbol{y}\} = \emptyset$.

Now we take $n = n + K(S)$ and split the general event from the second sequence in (A.1.4) according to the initial value of the process:

(A.1.6)

$$\big\{\mathrm{span}\,\big\{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_{n+K(S)}\big\} \subseteq S\big\} = \bigcup_{\boldsymbol{y}_1 \in \mathbb{Z}_+^p \times \{1\}} \{\mathrm{span}\,\big\{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_{n+K(S)}\big\} \subseteq S, \boldsymbol{Y}_1 = \boldsymbol{y}_1\}.$$

The individual events in the union can be transformed in the following way:

(A.1.7)
$$
\begin{aligned}
&\{\text{span}\left\{\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{n+K(S)}\right\} \subseteq S, \boldsymbol{Y}_1=\boldsymbol{y}_1\} = \{\boldsymbol{Y}_1\in S,\ldots,\boldsymbol{Y}_{n+K(S)}\in S, \boldsymbol{Y}_1=\boldsymbol{y}_1\}\\
&\quad = \{\boldsymbol{Y}_1\in S,\ldots,\boldsymbol{Y}_{K(S)}\in S, \boldsymbol{Y}_1=\boldsymbol{y}_1, \boldsymbol{Y}_{1+K(S)}\in S,\ldots,\boldsymbol{Y}_{n+K(S)}\in S\}\\
&\quad \subseteq \{\boldsymbol{Y}_1\in S,\ldots,\boldsymbol{Y}_{K(S)}\in S, \boldsymbol{Y}_1=\boldsymbol{y}_1, X_1=m(y_1,S), \boldsymbol{Y}_{2+K(S)}\in S,\ldots,\boldsymbol{Y}_{n+K(S)}\in S\}\\
&\quad = \{\boldsymbol{Y}_1\in S,\ldots,\boldsymbol{Y}_{K(S)}\in S, \boldsymbol{Y}_1=\boldsymbol{y}_1, \boldsymbol{Y}_2=\boldsymbol{y}_2, \boldsymbol{Y}_{2+K(S)}\in S,\ldots,\boldsymbol{Y}_{n+K(S)}\in S\}\\
&\quad \subseteq \{\boldsymbol{Y}_1\in S,\ldots,\boldsymbol{Y}_{K(S)}\in S,\\
&\qquad\quad \boldsymbol{Y}_1=\boldsymbol{y}_1, \boldsymbol{Y}_2=\boldsymbol{y}_2, X_2=m(y_2,S), \boldsymbol{Y}_{3+K(S)}\in S,\ldots,\boldsymbol{Y}_{n+K(S)}\in S\}\\
&\quad = \{\boldsymbol{Y}_1\in S,\ldots,\boldsymbol{Y}_{K(S)}\in S, \boldsymbol{Y}_1=\boldsymbol{y}_1, \boldsymbol{Y}_2=\boldsymbol{y}_2, \boldsymbol{Y}_3=\boldsymbol{y}_3, \boldsymbol{Y}_{3+K(S)}\in S,\ldots,\boldsymbol{Y}_{n+K(S)}\in S\}\\
&\quad \vdots\\
&\quad \subseteq \{\boldsymbol{Y}_1\in S,\ldots,\boldsymbol{Y}_{K(S)}\in S, \boldsymbol{Y}_1=\boldsymbol{y}_1, \boldsymbol{Y}_2=\boldsymbol{y}_2, \boldsymbol{Y}_3=\boldsymbol{y}_3,\ldots,\boldsymbol{Y}_n=\boldsymbol{y}_n\},
\end{aligned}
$$

where the sequence $(\boldsymbol{y}_i)_{i=1}^n$ is defined by the recursion

(A.1.8)
$$
\boldsymbol{y}_i = \begin{bmatrix} m(\boldsymbol{y}_{i-1},S) \\ y_{i-1}^{(1)} \\ \vdots \\ y_{i-1}^{(p-1)} \\ 1 \end{bmatrix}, \quad i=2,3,\ldots,n.
$$

We would like to represent the probability of the last event in (A.1.7) as a product of transition probabilities. For this we first need to determine whether the event is empty, and now we will give two necessary conditions on $\boldsymbol{y}_1$ for its nonemptiness. The first condition is, clearly, that all elements of the sequence defined in (A.1.8) fall into $\mathbb{Z}_+^p\times\{1\}$. We will not investigate this condition in any further detail, we only note that this imposes a deterministic condition on $\boldsymbol{y}_1$. Another deterministic condition is that $\boldsymbol{Y}_1\in S, \boldsymbol{Y}_2\in S,\ldots,\boldsymbol{Y}_{K(S)}\in S$ should all hold. Because the first $K(S)-1$ coefficients are all zero in any equation in (A.1.5) and $\boldsymbol{Y}_1$ contains all the components indexed $K(S)$ or greater in $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{K(S)}$, the validity of these inclusions is determined by $\boldsymbol{y}_1$ alone. This imposes the second (again, deterministic) condition on $\boldsymbol{y}_1$. If we denote the set of $\boldsymbol{y}_1$ which fulfill both these conditions by U_n, we have from (A.1.6) and (A.1.7),

$$
\left\{\text{span}\left\{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\ldots,\boldsymbol{Y}_{n+K(S)}\right\}\subseteq S\right\} \subseteq \bigcup_{\boldsymbol{y}_1\in U_n}\{\boldsymbol{Y}_1=\boldsymbol{y}_1, \boldsymbol{Y}_2=\boldsymbol{y}_2, \boldsymbol{Y}_3=\boldsymbol{y}_3,\ldots,\boldsymbol{Y}_n=\boldsymbol{y}_n\},
$$

hence by σ-subadditivity (U_n is clearly countable),

$$\begin{aligned}\mathbb{P}\left(\operatorname{span}\left\{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_{n+K(S)}\right\} \subseteq S\right) &\leqslant \sum_{\boldsymbol{y}_1 \in U_n} \mathbb{P}\left(\boldsymbol{Y}_1 = \boldsymbol{y}_1, \boldsymbol{Y}_2 = \boldsymbol{y}_2, \ldots, \boldsymbol{Y}_n = \boldsymbol{y}_n\right) \\ &= \sum_{\boldsymbol{y}_1 \in U_n} \mathbb{P}(\boldsymbol{Y}_1 = \boldsymbol{y}_1) p_{\boldsymbol{y}_1, \boldsymbol{y}_2} p_{\boldsymbol{y}_2, \boldsymbol{y}_3} \cdots p_{\boldsymbol{y}_{n-1}, \boldsymbol{y}_n},\end{aligned} \tag{A.1.9}$$

where $p_{\boldsymbol{u},\boldsymbol{v}}$ denotes the transition probability of $\boldsymbol{Y}$ from $\boldsymbol{u}$ to $\boldsymbol{v}$. Because the sets $(U_n)_{n\in\mathbb{N}}$ form a nonincreasing sequence (the second condition does not depend on n, and the first one become more restrictive as n increases), it is sufficient to show that for any sequence $(\boldsymbol{y}_i)_{i\in\mathbb{N}} \in (\mathbb{N}_0^p \times \{1\})^{\mathbb{N}}$ we have

$$\lim_{n\to\infty} p_{\boldsymbol{y}_1, \boldsymbol{y}_2} p_{\boldsymbol{y}_2, \boldsymbol{y}_3} \cdots p_{\boldsymbol{y}_{n-1}, \boldsymbol{y}_n} = 0, \tag{A.1.10}$$

and from this we will get (A.1.4). For the proof of (A.1.10) we will need to establish upper bounds for the transition probabilities. We will first consider the case when $\sigma > 0$, i.e., when the innovation distribution is nondegenerate.

Let us fix $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{N}_0^p \times \{1\}$ so that $v^{(2)} = u^{(1)}, v^{(3)} = u^{(2)}, \ldots, v^{(p)} = u^{(p-1)}, v^{(p+1)} = 1$. We would like to give an upper bound for $p_{\boldsymbol{u},\boldsymbol{v}}$. We have for every $i \in \mathbb{N}$ and any $m \in \mathbb{Z}_+$,

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{Y}_{i+1} = \boldsymbol{v} \big| \boldsymbol{Y}_i = \boldsymbol{u}, \sum_{j=1}^{p} \sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell} = m\right) &= \mathbb{P}\left(\varepsilon_i = v^{(1)} - m \big| \boldsymbol{Y}_i = \boldsymbol{u}, \sum_{j=1}^{p} \sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell} = m\right) \\ &= \mathbb{P}\left(\varepsilon_i = v^{(1)} - m\right).\end{aligned}$$

Applying the law of total probability we get

$$p_{\boldsymbol{u},\boldsymbol{v}} = \sum_{m\in\mathbb{Z}_+} \mathbb{P}(\varepsilon_i = v^{(1)} - m)\, \mathbb{P}\left(\sum_{j=1}^{p} \sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell} = m\right) \leqslant \max_{k\in\mathbb{Z}_+} \mathbb{P}(\varepsilon_i = k) < 1, \tag{A.1.11}$$

since the innovation distribution was nondegenerate. Therefore, if $\sigma > 0$, then we have a uniform upper bound on the transition probabilities, which implies (A.1.10) immediately.

The other case is if the innovation distribution is degenerate. First we note that in this case the innovation is equal to its expectation $\mu > 0$ almost surely, so that all components of $\boldsymbol{Y}_i$ are positive for $i \geqslant p+1$. According to the conditions, there is a coefficient $\alpha_q, q \in \{1, 2, \ldots, p\}$ such that $0 < \alpha_q < 1$. Similarly to the previous reasoning, if additionally we suppose that all components of $\boldsymbol{u}$ and $\boldsymbol{v}$ are greater or equal to μ, we have

$$\begin{aligned}
&\mathbb{P}\left(\boldsymbol{Y}_{i+1} = \boldsymbol{v} \middle| \boldsymbol{Y}_i = \boldsymbol{u}, \mu + \left(\sum_{\substack{j=1\\j\neq q}}^{p}\sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell}\right) + \sum_{\ell=1}^{u^{(q)}-1} \xi_{q,i,\ell} = m\right)\\
&= \mathbb{P}\left(\xi_{q,i,u^{(q)}} = v^{(1)} - m \middle| \boldsymbol{Y}_i = \boldsymbol{u}, \mu + \left(\sum_{\substack{j=1\\j\neq q}}^{p}\sum_{\ell=1}^{u^{(j)}} \xi_{j,i,\ell}\right) + \sum_{\ell=1}^{u^{(q)}-1} \xi_{q,i,\ell} = m\right)\\
&= \mathbb{P}\left(\xi_{q,i,u^{(q)}} = v^{(1)} - m\right).
\end{aligned}$$

Here we note that $\xi_{q,i,u^{(q)}}$ is a meaningful notation because $u^{(q)} \geqslant \mu$ and μ is a positive integer. Applying the law of total probability again, we have that

$$p_{\boldsymbol{u},\boldsymbol{v}} \leqslant \max(\alpha_q, 1-\alpha_q) < 1,$$

which again gives a uniform upper bound for the transition probabilities and yields (A.1.10). With this our proof is complete. □

It may be worth noting that Lemma A.1.1 imposes very weak conditions on the process—we only neglect the trivial case when all innovation and offspring distributions are degenerate. Also, the lemma does not require that the process start from zero—the initial distribution can be arbitrarily chosen on U. This gives us a chance to prove two important corollaries.

A.1.2 Corollary. *For an INAR(p) process under the alternative hypothesis satisfying the assumptions of Lemma A.1.1 both before and after the change, and* $\tau = \max(1, \lfloor n\rho \rfloor)$ *for some* $\rho > 0$ *constant, we have*

$$\mathbb{P}(\boldsymbol{Q}_n \textit{is singular}) \to 0.$$

Proof. To show this statement we only need to note that due to Lemma A.1.1 we have

$$\mathbb{P}(\boldsymbol{Q}_{\lfloor n\rho \rfloor} \text{is singular}) \to 0,$$

and clearly

$$\{\boldsymbol{Q}_n \text{is singular}\} \subseteq \{\boldsymbol{Q}_{\lfloor n\rho \rfloor} \text{is singular}\}$$

due to the reasoning at the beginning of the proof of Lemma A.1.1. □

A.1.3 Corollary. *Under the conditions of Theorem 2.7.1 both* $\boldsymbol{Q}'$ *and* $\boldsymbol{Q}''$ *are positive definite.*

Proof. First we prove for $\boldsymbol{Q}'$. We note that Lemma A.1.1 did not impose any conditions on the initial distribution of the process $\boldsymbol{Y}$, therefore we can start the process from its stationary distribution (the existence of which is a trivial corollary of the existence of such a distribution for $\boldsymbol{X}$ before the change). Now, the singularity of $\boldsymbol{Q}'$ is equivalent to the condition

$$\mathbb{P}\left(\begin{bmatrix}\widetilde{X}' \\ 1\end{bmatrix} \in S\right) < 1, \quad \forall S < \mathbb{R}^{p+1}.$$

Let us now suppose that the stationary distribution of $\boldsymbol{Y}$ is concentrated on a proper subspace $S < \mathbb{R}^{p+1}$. From (A.1.4), however, we conclude that the probability of the process remaining in S forever is zero. As the distribution of $\boldsymbol{Y}_n$ is the stationary distribution for every time n, this is an immediate contradiction. Therefore $\boldsymbol{Q}'$ is nonsingular, but since it is a covariance matrix, it is positive semidefinite, therefore it has to be positive definite. The proof is the same for $\boldsymbol{Q}''$. □

A.2 The conditional moments of M_k

We shall now derive several moments of M_k conditionally on $\mathcal{F}_{k-1}$ (this calculation is a reproduction of that in T. Szabó (2011a)). Let us write M_k in the form

$$M_k = \sum_{j=1}^{X_{k-1}} (\xi_{1,k,j} - \alpha_1) + \sum_{j=1}^{X_{k-2}} (\xi_{2,k,j} - \alpha_2) + \ldots + \sum_{j=1}^{X_{k-p}} (\xi_{p,k,j} - \alpha_p) + (\varepsilon_k - \mu).$$

All the terms on the right hand side have zero mean and are independent of each other conditionally on $\mathcal{F}_{k-1}$, therefore

$$\mathbb{E}(M_k^2 | \mathcal{F}_{k-1}) = \alpha_1(1-\alpha_1)X_{k-1} + \ldots + \alpha_p(1-\alpha_p)X_{k-p} + \sigma^2.$$

Similarly,

$$\begin{aligned}\mathbb{E}(M_k^4 | \mathcal{F}_{k-1}) = &\sum_{i=1}^{p} \mathbb{E}((\xi_{i,1,1} - \alpha_i)^4) X_{k-i} + 3 \sum_{i,j=1, i\neq j}^{p} \mathbb{E}((\xi_{i,1,1} - \alpha_i)^2 (\xi_{j,1,1} - \alpha_j)^2) X_i X_j \\ &+ 6 \sum_{i=1}^{p} \binom{X_i}{2} \mathbb{E}^2((\xi_{i,1,1} - \alpha_i)^2) \\ &+ 6 \sum_{i=1}^{p} X_{k-i}\, \mathbb{E}((\xi_{i,1,1} - \alpha_i)^2 (\varepsilon_1 - \mu)^2) + \mathbb{E}((\varepsilon_1 - \mu)^4),\end{aligned}$$

and, after substituting the expectations,

$$\mathbb{E}(M_k^4|\mathcal{F}_{k-1}) = \boldsymbol{\alpha}_4^\top X_{k-i} + 3 \sum_{i,j=1, i\neq j}^{p} \alpha_i(1-\alpha_i)\alpha_j(1-\alpha_j)X_iX_j$$
$$+ 6 \sum_{i=1}^{p} \binom{X_i}{2} \alpha_i^2(1-\alpha_i)^2 + 6\boldsymbol{\alpha}_2^\top \boldsymbol{X}_{k-1}\sigma^2 + \mathbb{E}((\varepsilon_1 - \mu)^4).$$

A.3 Strong approximation for the test process

If one is prepared to get more involved in approximation theory than strictly necessary for Theorem 2.5.1, the result found there can be improved considerably. As this is theoretically interesting, but doesn't change our tests at all, we will discuss it here in the Appendix. This formed the backbone of T. Szabó (2011a), and was indeed the author's first result on the problems considered in the present thesis. We will begin with a simple calculus result:

A.3.1 Proposition. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative numbers such that*

$$\frac{a_n}{n^\beta} \to 0 \quad \text{as } n \to \infty$$

for some $\beta > 0$. Then

$$\sup_{1\leqslant k\leqslant n} \frac{a_k}{n^\beta} \to 0 \quad \text{as } n \to \infty.$$

Proof. Based on the assumption we have

$$\forall \epsilon > 0 : \exists \nu(\epsilon) > 0 : \; \frac{a_n}{n^\beta} < \epsilon, \text{ if } n > \nu(\epsilon).$$

For a fixed $\epsilon > 0$ put

$$S(\epsilon) := \sup_{1\leqslant i\leqslant \nu(\epsilon)} a_i$$

and

$$\nu_2(\epsilon) := \max\left(\nu(\epsilon), \; \left(\frac{S(\epsilon)}{\epsilon}\right)^{1/\beta}\right).$$

We will conclude the proof by showing that if $n > \nu_2(\epsilon)$, then

$$\sup_{1\leqslant k\leqslant n} \frac{a_k}{n^\beta} < \epsilon. \tag{A.3.1}$$

Indeed, if $1 \leqslant k \leqslant \nu(\epsilon)$, then

$$\frac{a_k}{n^\beta} \leqslant \frac{S(\epsilon)}{\frac{S(\epsilon)}{\epsilon}} = \epsilon.$$

On the other hand, if $\nu(\epsilon) \leqslant k \leqslant n$, then

$$\frac{a_k}{n^\beta} \leqslant \frac{a_k}{k^\beta} \leqslant \epsilon,$$

which completes the proof. □

The starting point for strong approximation is the following theorem:

A.3.2 Theorem. (Eberlein, 1986) *Let $(\boldsymbol{Y}_k)_{k\geqslant 1}$ be a sequence of d-dimensional random vectors, $\boldsymbol{T}_k(\ell) := \boldsymbol{Y}_{\ell+1} + \cdots + \boldsymbol{Y}_{\ell+k}$ for $\ell \geqslant 0$, $k \geqslant 1$, and $\mathcal{G}_\ell$ the σ-algebra generated by the random vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\ell$. Assume*

(i) $\mathbb{E}(\boldsymbol{Y}_k) = 0$ *for all* $k \geqslant 1$.

(ii) *There exists $\delta > 0$ such that*

$$\sup_{\ell\geqslant 0} \|\mathbb{E}(\boldsymbol{T}_k(\ell) \,|\, \mathcal{G}_\ell)\|_1 = \mathrm{O}(k^{1/2-\delta}),$$

where $\|\cdot\|_1$ denotes the L_1-norm.

(iii) *There exists $\delta > 0$ such that*

$$\sup_{\ell\geqslant 0} \left\| \mathbb{E}\left\{\boldsymbol{T}_k(\ell)\boldsymbol{T}_k(\ell)^\top \,|\, \mathcal{G}_\ell\right\} - \mathbb{E}\left\{\boldsymbol{T}_k(\ell)\boldsymbol{T}_k(\ell)^\top\right\} \right\|_1 = \mathrm{O}(k^{1-\delta}),$$

where $\|X\|_1$ denotes the sum of the L_1-norms of the entries of a random matrix X.

(iv) *There exist $\delta > 0$ and a covariance matrix $\boldsymbol{\Sigma}$ such that*

$$\sup_{\ell\geqslant 0} \left\| k^{-1}\,\mathbb{E}\left\{\boldsymbol{T}_k(\ell)\boldsymbol{T}_k(\ell)^\top\right\} - \boldsymbol{\Sigma} \right\| = \mathrm{O}(k^{-\delta}).$$

(v) *There exists $\delta > 0$ such that*

$$\sup_{k\geqslant 1} \mathbb{E}(\|\boldsymbol{Y}_k\|^{2+\delta}) < \infty.$$

Then there exist $\kappa > 2$ and a d-dimensional standard Wiener process $(\boldsymbol{\mathcal{W}}(x))_{x\geqslant 0}$ such that, almost surely,

$$\sum_{k=1}^{\lfloor t\rfloor} \boldsymbol{Y}_k - \boldsymbol{\Sigma}^{1/2}\,\boldsymbol{\mathcal{W}}(t) = \mathrm{O}(t^{1/\kappa}) \quad \text{as } t \to \infty.$$

With the help of this, we can prove the following stronger version of Theorem 2.5.3:

A.3.3 Theorem. *Under condition* $\mathbf{C}_0$ *from Definition 2.2.1, there exists a* $p+1$*-dimensional standard Wiener process* $\big(\boldsymbol{\mathcal{W}}(x)\big)_{x\geqslant 0}$ *such that, with some* $\kappa > 2$,

$$\sum_{j=1}^{k} \boldsymbol{Z}_j - \boldsymbol{I}^{1/2}\boldsymbol{\mathcal{W}}(k) = o(k^{1/\kappa}) \quad \textit{a.s., as } k \to \infty.$$

Proof. We will set $\boldsymbol{Y}_k := \boldsymbol{Z}_k$ and show that the conditions for Theorem A.3.2 hold. Because $\boldsymbol{Z}_k$ are martingale differences, conditions (i) and (ii) (with $k = 1/2$) are fulfilled automatically.

For condition (iii) we first observe that $\boldsymbol{Z}_k$ are pairwise uncorrelated with respect to any $\mathcal{F}_\ell$ (again due to the law of iterated expectations) and $\mathcal{G}_\ell = \mathcal{F}_\ell,\ \ell \in \mathbb{N}$, hence

$$\mathbb{E}(\boldsymbol{T}_k(\ell)\boldsymbol{T}_k(\ell)^\top|\mathcal{G}_\ell) = \mathbb{E}(\boldsymbol{Z}_{\ell+1}\boldsymbol{Z}_{\ell+1}^\top|\mathcal{F}_\ell) + \ldots + \mathbb{E}(\boldsymbol{Z}_{\ell+k}\boldsymbol{Z}_{\ell+k}^\top|\mathcal{F}_\ell).$$

Let us introduce

$$\boldsymbol{B}_{\ell,i} := \mathbb{E}(\boldsymbol{Z}_{\ell+i}\boldsymbol{Z}_{\ell+i}^\top|\mathcal{F}_\ell) = \mathbb{E}\left(\left.\begin{bmatrix} M_{\ell+i}^2\boldsymbol{X}_{\ell+i-1}\boldsymbol{X}_{\ell+i-1}^\top & M_{\ell+i}^2\boldsymbol{X}_{\ell+i-1} \\ M_{\ell+i}^2\boldsymbol{X}_{\ell+i-1}^\top & M_{\ell+i}^2 \end{bmatrix}\right| \mathcal{F}_\ell\right)$$

Let us denote by $\boldsymbol{V}_{\ell,i}$ the column vector

$$\text{(A.3.2)} \qquad \boldsymbol{V}_{\ell,i} := \begin{bmatrix} \mathbb{E}(M_{\ell+i}^2|\mathcal{F}_\ell) \\ \mathbb{E}(M_{\ell+i}^2\boldsymbol{X}_{\ell+i-1}^{\otimes 2}|\mathcal{F}_{k-1}) \\ \mathbb{E}(M_{\ell+i}^2\boldsymbol{X}_{\ell+i-1}|\mathcal{F}_\ell) \\ \mathbb{E}(\boldsymbol{X}_{\ell+i-1}^{\otimes 3}|\mathcal{F}_\ell) \\ \mathbb{E}(\boldsymbol{X}_{\ell+i-1}^{\otimes 2}|\mathcal{F}_\ell) \\ \mathbb{E}(\boldsymbol{X}_{\ell+i-1}|\mathcal{F}_\ell) \end{bmatrix}$$

Clearly, all the entries of $\boldsymbol{B}_{\ell,i}$ are contained in $\boldsymbol{V}_{\ell,i}$. Therefore, if we show that

$$\sup_{k\geqslant 1}\sup_{\ell\geqslant 0}\left\|\sum_{i=1}^{k}\boldsymbol{V}_{\ell,i} - \sum_{i=1}^{k}\mathbb{E}(\boldsymbol{V}_{\ell,i})\right\|_1 < \infty,$$

then we can conclude that (iii) holds with $\delta = 1$. First we want to show that the following recursion applies:

$$\text{(A.3.3)} \qquad \boldsymbol{V}_{\ell,i+1} = \boldsymbol{R}\boldsymbol{V}_{\ell,i} + \boldsymbol{R}, i \geqslant 1$$

where $\boldsymbol{R}$ is a block upper triangular matrix containing only $0, \boldsymbol{A}, \boldsymbol{A}^{\otimes 2}, \boldsymbol{A}^{\otimes 3}$ in its main diagonal, and $\boldsymbol{R}$ is a column vector whose entries depend only on the moments of $\xi_i(1,1), 1 \leqslant i \leqslant p$ and $\varepsilon(1)$. Here $\boldsymbol{A}$ is the coefficient matrix defined in (2.2.3); however, $\boldsymbol{R}$ is not the same matrix as in that section. The initial condition is

$$\boldsymbol{V}_{\ell,1} = \mathbb{P}(\boldsymbol{X}_\ell),$$

where the entries of $\mathbb{P}(\boldsymbol{X}_\ell)$ are third-degree multivariate polynomials of the entries of $\boldsymbol{X}_\ell$. To verify (A.3.3), first note that the first 3 components of $\boldsymbol{V}_{\ell,i}$ can be expressed as a linear combination of the entries of the last 3 components, since these contain the conditional expectations of all possible three-factor products

$$X_{\ell+i-i_1}X_{\ell+i-i_2}X_{\ell+i-i_3}, \quad 1 \leqslant i_1, i_2, i_3 \leqslant p.$$

To prove this, we apply the law of iterated expectations:

$$\mathbb{E}(M^j_{\ell+i}|\mathcal{F}_\ell) = \mathbb{E}(\mathbb{E}(M^j_{\ell+i}|\mathcal{F}_{\ell+i-1})|\mathcal{F}_\ell), \quad j = 1, 2,$$

whence we can refer to Section A.2 and note that $\mathbb{E}(M^{j_1}_{\ell+i}|\mathcal{F}_{\ell+i-1})$ can be expressed by a linear combination of entries of $\boldsymbol{X}_{k-1}$ and a constant term. It remains to show that for all $1 \leqslant j \leqslant 4$, the conditional expectation $\mathbb{E}(\boldsymbol{X}^{\otimes j}_{\ell+i-1}|\mathcal{F}_\ell)$ can be expressed as a linear combination of the entries of $\mathbb{E}(\boldsymbol{X}^{\otimes j_1}_{\ell+i-2}|\mathcal{F}_\ell), 1 \leqslant j_1 \leqslant j$ and a constant term depending only on the moments of $\xi_i(1,1),\ 1 \leqslant i \leqslant p$ and $\varepsilon(1)$. Consider

$$\mathbb{E}(\boldsymbol{X}^{\otimes j}_{\ell+i-1}|\mathcal{F}_\ell) = \mathbb{E}((\boldsymbol{M}_{\ell+i-1} + A(\boldsymbol{X}_{\ell+i-2}) + \boldsymbol{\mu})^{\otimes j}|\mathcal{F}_\ell),$$

based on the regression equation (2.2.5), where

$$\boldsymbol{M}_k := M_k\mathbf{1}, \qquad \boldsymbol{\mu} := \mu\mathbf{1}. \tag{A.3.4}$$

The Kronecker product is not symmetric but linear in both factors, hence after expansion of $(\boldsymbol{M}_{\ell+i-1} + A(\boldsymbol{X}_{\ell+i-2}) + \boldsymbol{\mu})^{\otimes j}$ the resulting terms will be tensor products with factors $\boldsymbol{X}_{\ell+i-2}, \boldsymbol{M}_{\ell+i-1}, \boldsymbol{\mu}$ such that the sum of the exponents is j. One such term will be $\boldsymbol{X}^{\otimes j}_{\ell+i-2}$, which will account for $\boldsymbol{A}^{\otimes j}$ in the main diagonal of $\boldsymbol{R}$. The entries of the other terms will be products with factors $M_{\ell+i-1}$, μ and the entries of $\boldsymbol{X}_{\ell+i-2}$. The sum of the exponents should again be equal to j. For a typical product, we can put

$$\begin{aligned}
&\mathbb{E}(M^{j_1}_{\ell+i-1}\mu^{j_2}_\varepsilon X^{j_3}_{\ell+i-i_1}X^{j_4}_{\ell+i-i_2}X^{j_5}_{\ell+i-i_3}|\mathcal{F}_\ell) \\
&\qquad = \mathbb{E}(\mathbb{E}(M^{j_1}_{\ell+i-1}\mu^{j_2}_\varepsilon X^{j_3}_{\ell+i-i_1}X^{j_4}_{\ell+i-i_2}X^{j_5}_{\ell+i-i_3}|\mathcal{F}_{\ell+i-2})|\mathcal{F}_\ell),\ 2 \leqslant i_1, i_2, i_3 \leqslant p+1, \\
&\qquad\quad j_1 + j_2 + j_3 + j_4 + j_5 = j.
\end{aligned}$$

All the factors but $M_{\ell+i-1}$ are $\mathcal{F}_{\ell+i-2}$-measurable and so we can again apply the results of Section A.2 to show the desired statement. Note that in the terms appearing in $\mathbb{E}(M^{j_1}_{\ell+i-1}|\mathcal{F}_{\ell+i-2})$ the combined exponent of the X's is strictly smaller than j_1. Therefore there will be no added terms to $\boldsymbol{A}^{\otimes j}$ in the main diagonal of $\boldsymbol{R}$ in the rows corresponding to $\mathbb{E}(\boldsymbol{X}_{\ell+i-1}|\mathcal{F}_\ell)$.

Having proved (A.3.3) we can proceed with the proof of (iii). The recursion implies

$$\boldsymbol{V}_{\ell,i} = \sum_{j=2}^{i} \boldsymbol{R}^{i-j}\boldsymbol{R} + \boldsymbol{R}^{i-1}\,\mathbb{P}(\boldsymbol{X}_\ell), \quad i \geqslant 1 \tag{A.3.5}$$

The matrix $\boldsymbol{R}$ is block upper triangular, therefore its eigenvalues are the eigenvalues of the blocks in its main diagonal: $\boldsymbol{A}, \boldsymbol{A}^{\otimes 2}$ and $\boldsymbol{A}^{\otimes 3}$. The eigenvalues of these blocks are less than 1 (it is a well-known property of the Kronecker product that $\rho(\boldsymbol{A}\otimes\boldsymbol{B}) = \rho(\boldsymbol{A})\rho(\boldsymbol{B})$), thus we can conclude that $\rho(\boldsymbol{R}) < 1$ holds for the spectral radius. Therefore

$$\lim_{i\to\infty} \boldsymbol{V}_{0,i} = \sum_{j=0}^{\infty} (\boldsymbol{R})^j \boldsymbol{R} = (\boldsymbol{E}-\boldsymbol{R})^{-1}\boldsymbol{R} =: \boldsymbol{V}_0,$$

where $\boldsymbol{E}$ is an identity matrix of appropriate dimension. Because all the entries of $\boldsymbol{V}_0$ are finite, we conclude that

$$\lim_{k\to\infty} \mathbb{E}(\boldsymbol{X}_k^{\otimes\beta}) = \lim_{k\to\infty} \mathbb{E}(\boldsymbol{X}_k^{\otimes\beta}|\mathcal{F}_0) =: \boldsymbol{\eta}_\beta \tag{A.3.6}$$

exists for $\beta = 1,2,3$ and therefore it must also be the expectation $\mathbb{E}(\widetilde{\boldsymbol{X}}^{\otimes\beta})$, where $\widetilde{\boldsymbol{X}}$ has the unique stationary distribution of $\boldsymbol{X}_k$.

Now, (A.3.5) implies

$$\begin{aligned}\left\|\sum_{i=1}^{k} \boldsymbol{V}_{\ell,i} - \sum_{i=1}^{k} \mathbb{E}(\boldsymbol{V}_{\ell,i})\right\|_1 &= \left\|\left(\sum_{i=1}^{k} \boldsymbol{R}^{i-1}\right)(\mathbb{P}(\boldsymbol{X}_\ell) - \mathbb{E}(\mathbb{P}(\boldsymbol{X}_\ell)))\right\|_1 \\ &\leqslant \left\|\sum_{i=0}^{k-1} \boldsymbol{R}^i\right\|_1 \left\|\mathbb{P}(\boldsymbol{X}_\ell) - \mathbb{E}(\mathbb{P}(\boldsymbol{X}_\ell))\right\|_1 .\end{aligned}$$

Using (A.3.6) we can conclude

$$\left\|\mathbb{P}(\boldsymbol{X}_\ell) - \mathbb{E}(\mathbb{P}(\boldsymbol{X}_\ell))\right\|_1 \leqslant \left\|\mathbb{P}(\boldsymbol{X}_\ell) - \mathbb{E}(\mathbb{P}(\boldsymbol{X}_\ell))\right\|_2 \to \left\|\mathbb{P}\left(\widetilde{\boldsymbol{X}}\right) - \mathbb{E}\left(\mathbb{P}\left(\widetilde{\boldsymbol{X}}\right)\right)\right\|_2 < \infty,$$

as $\ell\to\infty$ where $\widetilde{\boldsymbol{X}}$ has the unique stationary distribution of $\boldsymbol{X}_k$. Therefore

$$\sup_{\ell\geqslant 0} \left\|\mathbb{P}(\boldsymbol{X}_\ell) - \mathbb{E}(\mathbb{P}(\boldsymbol{X}_\ell))\right\|_1 < \infty. \tag{A.3.7}$$

Because

$$\lim_{i\to\infty} \sum_{i=0}^{k-1} \boldsymbol{R}^i = \sum_{i=0}^{\infty} \boldsymbol{R}^i = (\boldsymbol{E}-\boldsymbol{R})^{-1}$$

exists, we have

$$\sup_{k\geqslant 1}\left\|\sum_{i=0}^{k-1}\boldsymbol{R}^i\right\|_1 < \infty$$

and together with (A.3.7) this implies

$$\sup_{k\geqslant 1}\sup_{\ell\geqslant 0}\left\|\sum_{i=1}^{k}\boldsymbol{V}_{\ell,i}-\sum_{i=1}^{k}\mathbb{E}(\boldsymbol{V}_{\ell,i})\right\|_1 < \infty,$$

which completes the proof of (iii).

Now we show that (iv) is satisfied with $\boldsymbol{\Sigma}=\boldsymbol{I}$ and $\delta=1$. Similarly to the proof of (iii), we first notice that

$$\mathbb{E}\left(\boldsymbol{T}_k(\ell)\boldsymbol{T}_k(\ell)^\top\right)=\mathbb{E}\left(\boldsymbol{Z}_{\ell+1}\boldsymbol{Z}_{\ell+1}^\top\right)+\ldots+\mathbb{E}\left(\boldsymbol{Z}_{\ell+k}\boldsymbol{Z}_{\ell+k}^\top\right).$$

Now we take, from (2.5.1),

$$\boldsymbol{I}_n=\sum_{k=1}^{n}\mathbb{E}\left(\boldsymbol{Z}_k\boldsymbol{Z}_k^\top|\mathcal{F}_{k-1}\right),\quad \mathbb{E}(\boldsymbol{I}_n)=\sum_{k=1}^{n}\mathbb{E}\left(\boldsymbol{Z}_k\boldsymbol{Z}_k^\top\right),$$

so that

$$\mathbb{E}\left(\boldsymbol{T}_k(\ell)\boldsymbol{T}_k(\ell)^\top\right)=\mathbb{E}(\boldsymbol{I}_{\ell+k})-\mathbb{E}(\boldsymbol{I}_k)$$

Let us consider the form of $\boldsymbol{I}_n$ and $\boldsymbol{I}$. The entries of $\mathbb{E}(\boldsymbol{I}_n)$ are entries from $\sum_{k=1}^n\mathbb{E}(\boldsymbol{X}_{k-1})$, $\sum_{k=1}^n\mathbb{E}(\boldsymbol{X}_{k-1}^{\otimes 2})$ and $\sum_{k=1}^n\mathbb{E}(\boldsymbol{X}_{n-1}^{\otimes 3})$ multiplied by constants. In $\boldsymbol{I}$ these moments are replaced by the respective moments of the stationary distribution, but the multiplicating constants remain the same. Hence, it will be enough to show

$$\sup_{\ell\geqslant 0}\sum_{j=1}^{k}\left\|\mathbb{E}\,\boldsymbol{X}_{\ell+j-1}^{\otimes\beta}-\boldsymbol{\eta}_\beta\right\|=\mathrm{O}(1)\quad\text{as } k\to\infty,\ \beta=1,2,3.$$

For this, it is sufficient that, with $m:=\rho(\boldsymbol{R})$,

$$\mathbb{E}(\boldsymbol{X}_i^{\otimes\beta})=\boldsymbol{\eta}_\beta+\mathrm{O}(m^i)\quad\text{as } i\to\infty,\ \beta=1,2,3, \tag{A.3.8}$$

because then, for some $K>0$,

$$\sup_{\ell\geqslant 0}\sum_{j=1}^{k}\left\|\mathbb{E}\,\boldsymbol{X}_{\ell+j-1}^{\otimes\beta}-\boldsymbol{\eta}_\beta\right\|\leqslant\sup_{\ell\geqslant 0}\sum_{j=1}^{k}Km^{\ell+j-1}\leqslant K\sum_{j=1}^{\infty}m^{j-1}=\frac{K}{1-m},$$

since $m < 1$. We can prove (A.3.8) by considering

$$\begin{aligned}\|\boldsymbol{V}_{0,i} - \boldsymbol{V}_0\| &= \left\| \boldsymbol{R}^{i-1}\,\mathbb{P}(\boldsymbol{X}_0) + \sum_{j=i-1}^{\infty} \boldsymbol{R}^j \boldsymbol{R} \right\| \\ &\leqslant \left\|\boldsymbol{R}^{i-1}\right\| \left(\|\mathbb{P}(\boldsymbol{X}_0)\| + \left\| \sum_{j=0}^{\infty} \boldsymbol{R}^j \right\| \|\boldsymbol{R}\| \right) = \mathrm{O}(m^i),\end{aligned}$$

because the second factor is constant, and by a well-known result in matrix analysis

$$\lim_{i\to\infty} \left\|\boldsymbol{R}^i\right\|^{1/i} = \rho(\boldsymbol{R}) = m.$$

For (v) we consider

$$\begin{aligned}\mathbb{E}(\|\boldsymbol{Z}_k\|^4 \,|\mathcal{F}_{k-1}) &= \mathbb{E}((M_k^2(X_{k-1}^2 + \ldots + X_{k-p}^2 + 1))^2 |\mathcal{F}_{k-1}) \\ &= (X_{k-1}^2 + \ldots + X_{k-p}^2 + 1)^2\,\mathbb{E}(M_k^4|\mathcal{F}_{k-1}) = \mathbb{P}(\boldsymbol{X}_{k-1}),\end{aligned}$$

where $\mathbb{P}(\boldsymbol{X}_{k-1})$ is a degree six polynomial of the entries of $\boldsymbol{X}_{k-1}$ (see A.2). Because $\mathbb{E}(\varepsilon_1^6) < \infty$, we have

$$\mathbb{E}(\mathbb{E}(\|\boldsymbol{Z}_k\|^4 \,|\mathcal{F}_{k-1})) \leqslant \mathbb{E}(\mathbb{P}(\boldsymbol{X}_{k-1})) \to \mathbb{E}(\mathbb{P}(\widetilde{\boldsymbol{X}})) < \infty$$

because of (A.3.8). $\square$

The following lemma provides the asymptotics for the partial sums of $\boldsymbol{X}_k$ and $\boldsymbol{X}_k^{\otimes 2}$.

A.3.4 Lemma. *If the assumptions of Theorem A.3.5 hold, then the sequence of random step processes*

$$\boldsymbol{\mathcal{X}}_n(t) := \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor} \begin{bmatrix} \boldsymbol{X}_k^{\otimes 2} \\ \boldsymbol{X}_k \end{bmatrix}, \quad t \in [0,1]$$

converges in distribution on the Skorokhod space $D([0,1])$:

$$\boldsymbol{\mathcal{X}}_n \xrightarrow{\mathcal{D}} \boldsymbol{\mathcal{X}},$$

where

$$\boldsymbol{\mathcal{X}}(t) := t \begin{bmatrix} \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_1 \end{bmatrix}$$

Proof. We apply the multidimensional martingale central limit theorem (see Theorem 3.3.4) for the sequences $(\boldsymbol{U}_{n,k}, \mathcal{F}_k)_{1\leqslant k\leqslant n}, n \geqslant 1$, where

$$\boldsymbol{U}_{n,k} := \frac{1}{n} \begin{bmatrix} \boldsymbol{X}_k^{\otimes 2} \\ \boldsymbol{X}_k \end{bmatrix}.$$

The ergodic theorem implies

$$\frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}(\boldsymbol{X}_k|\mathcal{F}_{k-1}) = \frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor}(\boldsymbol{A}\boldsymbol{X}_{k-1}+\boldsymbol{\mu}) \xrightarrow{\text{a.s.}} t(\boldsymbol{A}\boldsymbol{\eta}_1+\boldsymbol{\mu}) = t\boldsymbol{\eta}_1, t\in[0,1].$$

Similarly,

$$\frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}(\boldsymbol{X}_k^{\otimes 2}|\mathcal{F}_{k-1}) \xrightarrow{\text{a.s.}} t\boldsymbol{\eta}_2, \quad t\in[0,1]$$

Therefore, the asymptotic expectation is

$$\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}(\boldsymbol{U}_{n,k}|\mathcal{F}_{k-1}) \xrightarrow{\text{a.s.}} t\begin{bmatrix}\boldsymbol{\eta}_2\\ \boldsymbol{\eta}_1\end{bmatrix}.$$

The asymptotic covariance matrix is 0, because

$$\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}(\boldsymbol{U}_{n,k}\boldsymbol{U}_{n,k}^\top|\mathcal{F}_{k-1}) = \frac{1}{n^2}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_k^{\otimes 2}\\ \boldsymbol{X}_k\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_k^{\otimes 2}\\ \boldsymbol{X}_k\end{bmatrix}^\top \middle| \mathcal{F}_{k-1}\right) \xrightarrow{\text{a.s.}} 0, \quad t\in[0,1],$$

because

$$\frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}\left(\begin{bmatrix}\boldsymbol{X}_k^{\otimes 2}\\ \boldsymbol{X}_k\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_k^{\otimes 2}\\ \boldsymbol{X}_k\end{bmatrix}^\top \middle| \mathcal{F}_{k-1}\right) \xrightarrow{\text{a.s.}} \mathbb{E}\left(\begin{bmatrix}\widetilde{\boldsymbol{X}}^{\otimes 2}\\ \widetilde{\boldsymbol{X}}\end{bmatrix}\begin{bmatrix}\widetilde{\boldsymbol{X}}^{\otimes 2}\\ \widetilde{\boldsymbol{X}}\end{bmatrix}^\top\right), \quad t\in[0,1],$$

which is a finite quantity, as the entries of the matrix are all contained in $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4$, which are all finite. Furthermore, we have

$$\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}(\|\boldsymbol{U}_{n,k}\|^2|\mathcal{F}_{k-1}) = \frac{1}{n^2}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}(\|\boldsymbol{X}_k^{\otimes 2}\|^2 + \|\boldsymbol{X}_k\|^2|\mathcal{F}_{k-1}) \xrightarrow{\text{a.s.}} 0, \quad t\in[0,1],$$

because

$$\frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor}\mathbb{E}(\|\boldsymbol{X}_k^{\otimes 2}\|^2 + \|\boldsymbol{X}_k\|^2|\mathcal{F}_{k-1}) \xrightarrow{\text{a.s.}} \mathbb{E}\left(\left\|\widetilde{\boldsymbol{X}}^{\otimes 2}\right\|^2\right) + \mathbb{E}\left(\left\|\widetilde{\boldsymbol{X}}\right\|^2\right),$$

This proves the conditional Lindeberg condition. □

Now we are ready to state the strong approximation counterpart of Theorem 2.5.1.

A.3.5 Theorem. *Under* $\mathbf{C}_0$ *we can define a sequence of* $p+1$*-dimensional standard Brownian bridges* $(\boldsymbol{\mathcal{B}}_n(t))_{0\leqslant t\leqslant 1}$, $n\geqslant 1$, *such that*

$$\sup_{0\leqslant t\leqslant 1}\left\|\widehat{\boldsymbol{\mathcal{M}}}_n(t) - \boldsymbol{\mathcal{B}}_n(t)\right\| = o_{\mathbb{P}}(1) \qquad as \quad n\to\infty.$$

We can reuse the first few steps from the proof of Theorem 2.5.1, up until (2.5.7). Continuing from there, $\widehat{\boldsymbol{\mathcal{M}}}_n(t)$ takes the form

$$\widehat{\boldsymbol{\mathcal{M}}}_n(t) = \widehat{\boldsymbol{I}}_n^{-1/2} \left(\sum_{k=1}^{\lfloor nt \rfloor} \boldsymbol{Z}_k - \boldsymbol{\mathcal{A}}_n(t) \sum_{k=1}^{n} \boldsymbol{Z}_k \right)$$

with

$$\boldsymbol{\mathcal{A}}_n(t) := \sum_{k=1}^{\lfloor nt \rfloor} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \left(\sum_{k=1}^{n} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top \right)^{-1},$$

Let us now introduce the process

$$\boldsymbol{\mathcal{M}}_n(t) := n^{-1/2} \boldsymbol{I}^{-1/2} \left(\sum_{k=1}^{\lfloor nt \rfloor} \boldsymbol{Z}_k - t \sum_{k=1}^{n} \boldsymbol{Z}_k \right).$$

We will first show that there is a sequence $(\boldsymbol{\mathcal{B}}_n(t)_{0 \leqslant t \leqslant 1}), n \geqslant 1$ of $p+1$-dimensional standard Brownian bridges such that

$$\sup_{0 \leqslant t \leqslant 1} \|\boldsymbol{\mathcal{M}}_n(t) - \boldsymbol{\mathcal{B}}_n(t)\| = \mathrm{o}_{\mathbb{P}}(1) \quad \text{as } n \to \infty. \tag{A.3.9}$$

Let $\boldsymbol{\mathcal{W}}$ the Wiener process provided by Theorem A.3.3. Putting

$$w_n = \left\| \boldsymbol{I}^{-1/2} \sum_{k=1}^{n} \boldsymbol{Z}_k - \boldsymbol{\mathcal{W}}(n) \right\|$$

we have that $w_n = \mathrm{o}(n^{1/\kappa})$ on an event of probability 1. Now note that

$$\sup_{0 \leqslant t \leqslant 1} \left\| \boldsymbol{I}^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \boldsymbol{Z}_k - \boldsymbol{\mathcal{W}}(\lfloor nt \rfloor) \right\| = \max_{0 \leqslant k \leqslant n} w_k.$$

Using Proposition A.3.1 we have

$$\sup_{0 \leqslant t \leqslant 1} \left\| \boldsymbol{I}^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \boldsymbol{Z}_k - \boldsymbol{\mathcal{W}}(\lfloor nt \rfloor) \right\| = \mathrm{o}(n^{1/\kappa}) \quad \text{a.s.}$$

Consequently,

$$\sup_{0 \leqslant t \leqslant 1} \left\| n^{-1/2} \boldsymbol{I}^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \boldsymbol{Z}_k - n^{-1/2} \boldsymbol{\mathcal{W}}(\lfloor nt \rfloor) \right\| = \mathrm{o}\left(\frac{n^{1/\kappa}}{n^{1/2}} \right) = \mathrm{o}(1) \quad \text{a.s.}$$

and similarly

$$\sup_{0 \leqslant t \leqslant 1} \left\| n^{-1/2} \boldsymbol{I}^{-1/2} t \sum_{k=1}^{n} \boldsymbol{Z}_k - n^{-1/2} t \boldsymbol{\mathcal{W}}(n) \right\| = \mathrm{o}(1) \quad \text{a.s.},$$

because $\kappa > 2$. By the triangle inequality we conclude that

$$\left\| \boldsymbol{\mathcal{M}}_n(t) - n^{-1/2}(\boldsymbol{\mathcal{W}}(\lfloor nt \rfloor) - t\boldsymbol{\mathcal{W}}(n)) \right\| = \mathrm{o}(1) \quad \text{a.s.}$$

The process $\boldsymbol{\mathcal{B}}_n(t)$ will be defined by

$$n^{-1/2}(\boldsymbol{\mathcal{W}}(nt) - t\boldsymbol{\mathcal{W}}(n)),$$

which is obviously a Brownian bridge, since $n^{-1/2}\boldsymbol{\mathcal{W}}(nt)$ is a standard Wiener process. It remains to show that

$$\begin{aligned} &\sup_{0\leqslant t\leqslant 1} \left\| n^{-1/2}(\boldsymbol{\mathcal{W}}(\lfloor nt \rfloor) - t\boldsymbol{\mathcal{W}}(n)) - n^{-1/2}(\boldsymbol{\mathcal{W}}(nt) - t\boldsymbol{\mathcal{W}}(n)) \right\| \\ = &\sup_{0\leqslant t\leqslant 1} \left\| n^{-1/2}(\boldsymbol{\mathcal{W}}(\lfloor nt \rfloor) - \boldsymbol{\mathcal{W}}(nt)) \right\| = \mathrm{o}_{\mathbb{P}}(1). \end{aligned}$$

Because the components of $\boldsymbol{\mathcal{W}}(t)$ are independent, it suffices to show that

$$\text{(A.3.10)} \qquad \sup_{0\leqslant t\leqslant 1} \left| n^{-1/2}(\mathcal{W}^{(1)}(\lfloor nt \rfloor) - \mathcal{W}^{(1)}(nt)) \right| = \mathrm{o}_{\mathbb{P}}(1),$$

where $\mathcal{W}^{(1)}(t)$ is the first component of $\boldsymbol{\mathcal{W}}(t)$, and is a standard Wiener process. For any $\epsilon > 0$ we have

$$\begin{aligned} &\mathbb{P}\left(\sup_{0\leqslant t\leqslant 1} \left| n^{-1/2} \left(\mathcal{W}^{(1)}(\lfloor nt \rfloor) - \mathcal{W}^{(1)}(nt) \right) \right| > \epsilon \right) \\ &\qquad \leqslant \sum_{k=1}^{n} \mathbb{P}\left(\sup_{0\leqslant u\leqslant 1} \left| \mathcal{W}^{(1)}(k-1+u) - \mathcal{W}^{(1)}(k-1) \right| > \epsilon^{1/2} \right) \\ &\qquad \leqslant 4n \left(1 - \Phi\left(\epsilon n^{1/2} \right) \right) \end{aligned}$$

by Csörgő (2010, Proposition 51.1). By Csörgő (2010, Lemma 34.2) we have

$$4n \left(1 - \Phi\left(\epsilon n^{1/2} \right) \right) \leqslant \frac{n}{\sqrt{2\pi}\epsilon n^{1/2}} e^{-\frac{\epsilon^2 n}{2}} \to 0,$$

which proves (A.3.10) and thus, (A.3.9). Now we only need to show

$$\text{(A.3.11)} \qquad \sup_{0\leqslant t\leqslant 1} \left\| \widehat{\boldsymbol{\mathcal{M}}}_n(t) - \boldsymbol{\mathcal{M}}_n(t) \right\| = \mathrm{o}_{\mathbb{P}}(1) \quad \mathrm{as} n \to \infty.$$

We apply the triangle inequality:

$$\left\|\widehat{\boldsymbol{\mathcal{M}}}_n(t)-\boldsymbol{\mathcal{M}}_n(t)\right\| \leqslant \left\|\widehat{\boldsymbol{I}}_n^{-1/2}-n^{-1/2}\boldsymbol{I}^{-1/2}\right\| \left\|\sum_{k=1}^{\lfloor nt\rfloor}\boldsymbol{Z}_k - t\sum_{k=1}^{n}\boldsymbol{Z}_k\right\|$$
$$+\left\|\widehat{\boldsymbol{I}}_n\right\| \left\|\boldsymbol{\mathcal{A}}_n(t)-t\boldsymbol{E}\right\| \left\|\sum_{k=1}^{n}\boldsymbol{Z}_k\right\|.$$

We have $n^{-1}\widehat{\boldsymbol{I}}_n \to \boldsymbol{I}$ a.s., hence

$$\left\|\widehat{\boldsymbol{I}}_n^{-1/2}-\boldsymbol{I}^{-1/2}n^{-1/2}\right\| = \mathrm{o}(n^{-1/2}) \quad \text{a.s.}$$

and

$$\left\|\widehat{\boldsymbol{I}}_n\right\| = \mathrm{O}(n^{-1/2}) \quad \text{a.s.}$$

Using (A.3.9) and the definition of $\boldsymbol{\mathcal{M}}_n(t)$ we have

$$\sup_{0\leqslant t\leqslant 1}\left\|\sum_{k=1}^{\lfloor nt\rfloor}\boldsymbol{Z}_k - t\sum_{k=1}^{n}\boldsymbol{Z}_k\right\| \leqslant n^{1/2}\left\|\boldsymbol{I}^{-1/2}\right\|\left(\sup_{0\leqslant t\leqslant 1}\left\|\boldsymbol{\mathcal{M}}_n(t)-\boldsymbol{\mathcal{B}}_n(t)\right\| + \sup_{0\leqslant t\leqslant 1}\left\|\boldsymbol{\mathcal{B}}_n(t)\right\|\right)$$
$$= \mathrm{O}_{\mathbb{P}}(n^{1/2}).$$

Here we have used the fact that the processes $\boldsymbol{\mathcal{B}}_n(t)$ are identically distributed, therefore $\sup_{0\leqslant t\leqslant 1}\|\boldsymbol{\mathcal{B}}_n(t)\| = \mathrm{O}_{\mathbb{P}}(1)$. According to Theorem A.3.3, we have

$$\left\|\sum_{k=1}^{n}\boldsymbol{Z}_k\right\| \leqslant \left\|\sum_{k=1}^{n}\boldsymbol{Z}_k - \boldsymbol{I}^{1/2}\boldsymbol{\mathcal{W}}(n)\right\| + \left\|\boldsymbol{I}^{1/2}\boldsymbol{\mathcal{W}}(n)\right\| = \mathrm{O}_{\mathbb{P}}(n^{1/2}),$$

because $\left\|\boldsymbol{I}^{1/2}\boldsymbol{\mathcal{W}}(n)\right\| = \mathrm{O}_{\mathbb{P}}(n^{1/2})$ due to the well-known growth rate of the standard Wiener process.

Finally, because the supremum is a continuous functional on the space $C[0,1]$, we have by the continuous mapping theorem and Lemma A.3.4

$$\sup_{0\leqslant t\leqslant 1}\|\boldsymbol{\mathcal{A}}_n(t)-t\boldsymbol{E}\| = \mathrm{o}_{\mathbb{P}}(1).$$

We have shown (A.3.11), and this completes our proof. □

Appendix B

Summary

The thesis is concerned with providing statistical methods for detecting change in the parameters of a stochastic process. This is generally a longstanding problem in time series analysis (Csörgő and Horváth, 1997), but investigating it for branching processes has received less attention.

The basic setup for change detection will be the following:

1. We will take a vector-valued process $\boldsymbol{X}_t$, indexed either by the natural numbers or the nonnegative real numbers and take a sample of it on the interval $0 \leqslant t \leqslant T$.

2. We will choose a parameter $\boldsymbol{\theta}_t$ governing the dynamics of the process. The main question will be whether this parameter is constant in t, or, formally, we would like to test

$$\mathrm{H}_0 : \ \exists \boldsymbol{\theta} : \ \boldsymbol{\theta}_t = \boldsymbol{\theta}, \quad t \in [0, T]$$

against the alternative hypothesis

$$\mathrm{H}_\mathrm{A} : \exists \rho \in (0,1) : \ \boldsymbol{\theta}_t = \boldsymbol{\theta}', \ t \in [0, \rho T) \ \text{ and } \ \boldsymbol{\theta}_t = \boldsymbol{\theta}'', \ t \in [\rho T, T]$$

for some $\boldsymbol{\theta}' \neq \boldsymbol{\theta}''$. An important additional condition will be for stability: $\boldsymbol{\theta}$, $\boldsymbol{\theta}'$, $\boldsymbol{\theta}''$ have to be such that $\boldsymbol{X}$ have a unique stationary distribution under H_0, and both parts of the process (before and after the change) have a unique stationary distribution under H_A.

3. We will find an appropriate vector-valued function f such that

$$\boldsymbol{M}_t := \boldsymbol{X}_t - \boldsymbol{X}_0 - \int_0^t f(\boldsymbol{\theta}_s; \boldsymbol{X}_{s-}) \, \mathrm{d}s$$

will be a martingale. Here $\boldsymbol{X}_{s-}$, a slightly informal notation, means $\boldsymbol{X}_s$ for continuous s and $\boldsymbol{X}_{s-1}$ for discrete s. Similarly, the integral is simply a sum for discrete s.

4. Assuming $\boldsymbol{\theta}_t = \boldsymbol{\theta}$ for all t, we will estimate $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}_T$ based on the conditional least squares (CLS) method of Klimko and Nelson (1978).

5. We will replace $\boldsymbol{\theta}_t$ with $\widehat{\boldsymbol{\theta}}_T$ in the definition of $\boldsymbol{M}_t$ to obtain $\widehat{\boldsymbol{M}}_t^{(T)}$.

6. We will prove that if $\boldsymbol{\theta}_t$ is constant in t, then

$$\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)} := \widehat{\boldsymbol{I}}_T^{-1/2} \widehat{\boldsymbol{M}}_{uT}^{(T)}, \quad u \in [0,1]$$

converges in distribution to a Brownian bridge on $[0,1]$, for some random normalizing matrix $\widehat{\boldsymbol{I}}_T$, which is calculable from the sample.

7. Consequently, we will construct tests for the change in $\boldsymbol{\theta}$, using the supremum or infimum of $\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)}$ as a test statistic (based on the direction of change).

8. We will prove that if there is a single change in $\boldsymbol{\theta}_t$ on $[0,T]$, then the test statistic will tend to infinity stochastically as $T \to \infty$.

9. We will prove that the *arg max*, or *arg min*, of $\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)}$ is a good estimator of the change point in $\boldsymbol{\theta}_t$.

The INAR(p) process

In Chapter 2, based on Pap and Szabó (2013), we will prove results for the integer valued autoregressive process of order p (INAR(p)), defined by:

$$X_k = \alpha_1 \circ X_{k-1} + \cdots + \alpha_p \circ X_{k-p} + \varepsilon_k,\ k = 1, 2, \ldots,$$

where the ε_k are i.i.d nonnegative integer-valued random variables with mean μ, and $\circ$ is the binomial thinning operator: for a random nonnegative integer-valued random variable Y and $\alpha \in (0,1)$, $\alpha \circ Y$ denotes the sum of Y i.i.d Bernoulli random variables with mean α, also independent of Y. This process was first proposed by Alzaid and Al-Osh (1987) for $p = 1$ and Du and Li (1991) for higher p values.

The parameter vector is

$$\boldsymbol{\theta} := \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ \mu \end{bmatrix},$$

with the stability condition

$$\alpha_1 + \ldots + \alpha_p < 1.$$

We are able to use the standard conditional least squares estimates from (2.4.1) and classical martingale theory to create a test process that is shown in Theorem 2.5.1 to converge to a $p+1$-dimensional Brownian bridge. Based on this, we can construct the tests, as described in subsection 2.5.1.

The weak consistence of the test is established in Theorem 2.7.1. The proof uses a decomposition of the test process as given in (2.7.2), and Lemma 2.9.1 in order to estimate the suprema of the negligible terms in the decomposition. It turns out that essentially the same tools can be applied to prove the asymptotic properties of the change-point estimator, given in Theorem 2.8.1.

The Cox–Ingersoll–Ross process

Chapter 4, based on Pap and Szabó (2016), is about change detection in the Cox–Ingersoll–Ross (CIR) process:

$$\mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma\sqrt{Y_t}\,\mathrm{d}W_t,\ t \in \mathbb{R}_+,$$

where $a, b, \sigma > 0$ and $(W_t)_{t\geqslant 0}$ is a standard Wiener process. The constraint on a and b will be the stability condition itself.

This process was first investigated by Feller (1951), proposed as a short-term interest-rate model by Cox et al. (1985), and became one of the most widespread “short rate” models in financial mathematics. Inevitably, therefore, describing its statistical properties is of great importance and has received considerable interest.

Our parameter vector will be

$$\boldsymbol{\theta} := \begin{bmatrix} a \\ b \end{bmatrix}$$

It turns out that the theorems and their proofs can be constructed along the same lines as for the INAR(p) process. The estimates given in (4.1.1) are not the usual ones, but their structure is similar to the CLS estimates in the INAR(p) case, and the martingale in (4.2.1) is also similar. The main result under the null hypothesis is Theorem 4.2.1. Its proof depends on the same apparatus as Theorem 2.5.1. Also, it is apparent that the statements of Theorems 4.5.1 and 4.6.1 are very similar to Theorems 2.7.1 and 2.8.1, respectively. This is reflected in their proofs; however, some of the steps require more advanced tools. In particular, we believe Lemma 4.7.2 (replacing Lemma 2.9.1) to be a new result, and the proof of Lemma 4.7.7 is significantly more involved than that of Lemma 2.9.7.

Parameter estimation for the Heston model

In Chapter 5, based on Barczy et al. (2016), we propose conditional least squares estimates (CLSE's) for the Heston model, which is the solution of a two-dimensional stochastic differential equation:

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma_1\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (\alpha - \beta Y_t)\,\mathrm{d}t + \sigma_2\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big), \end{cases} \qquad t \geqslant 0,$$

where $a > 0$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$, and $(W_t, B_t)_{t\geqslant 0}$ is a 2-dimensional standard Wiener process. It is immediately apparent that Y is just the Cox–Ingersoll–Ross process introduced before. The stability condition is $b > 0$ here as well, similarly to the CIR process.

Introducing CLSE's for (a, b, α, β) based on discrete time observations turns out to be impractical, as the conditional means, and consequently, the resulting partial derivatives depend on the parameters in a complicated manner. Therefore we transform the parameter space, as defined in (5.1.3), and derive CLSE's for the transformed parameter vector, which will result in linear partial derivatives, given in (5.1.4). We prove strong consistence and asymptotic normality in Theorem 5.2.1, using the same tools of martingale theory as for the CIR process (collected in Chapter 3) – however, the calculations turn out to be much more cumbersome than in Chapter 4. Applying the inverse transformation to the CLSE's leads to estimates for the original parameters. These are given in (5.4.3), and their strong consistence and asymptotic normality is proven in Theorem 5.4.2, based on Theorem 5.2.1 and the so-called delta method.

Appendix C

Összefoglaló

Az értekezés célja, hogy olyan módszereket adjon, amellyel változást észlelhetünk egy sztochasztikus folyamat paramétereiben. Ez általánosságban véve egy régóta vizsgált probléma az idősoranalízisben (Csörgő és Horváth, 1997), de az elágazó folyamatok területén ezidáig kevesebb figyelmet kapott.

Módszerünk a következő lépésekből fog állni:

1. Veszünk egy vektorértékű $\boldsymbol{X}_t$ folyamatot, amelyet vagy a természetes, vagy a nemnegatív valós számokkal indexelünk, és mintát veszünk a folyamatból a $0 \leqslant t \leqslant T$ intervallumon.

2. Választunk egy $\boldsymbol{\theta}_t$ paramétert, amely a folyamat dinamikáját irányítja. A fő kérdés az lesz, hogy ez a paraméter t-ben állandó-e, vagyis tesztelni szeretnénk a

$$\mathrm{H}_0: \ \exists \boldsymbol{\theta}: \ \boldsymbol{\theta}_t = \boldsymbol{\theta}, \quad t \in [0, T]$$

nullhipotézist a

$$\mathrm{H}_\mathrm{A}: \exists \rho \in (0,1), \boldsymbol{\theta}' \neq \boldsymbol{\theta}'': \ \boldsymbol{\theta}_t = \boldsymbol{\theta}', \ t \in [0, \rho T) \text{ és } \boldsymbol{\theta}_t = \boldsymbol{\theta}'', \ t \in [\rho T, T]$$

alternatív hipotézissel szemben. Fontos további feltétel lesz a stabilitás: $\boldsymbol{\theta}$, $\boldsymbol{\theta}'$, $\boldsymbol{\theta}''$ olyanok kell legyenek, hogy H_0 mellett $\boldsymbol{X}$-nek legyen egyértelmű stacionárius eloszlása, H_A mellett pedig ez a folyamat változás előtti és változás utáni részére is teljesüljön.

3. Keresünk egy vektorértékű f függvényt, amelyre

$$\boldsymbol{M}_t := \boldsymbol{X}_t - \boldsymbol{X}_0 - \int_0^t f(\boldsymbol{\theta}_s; \boldsymbol{X}_{s-}) \, \mathrm{d}s$$

martingál lesz. Itt $\boldsymbol{X}_{s-}$ egyszerűen $\boldsymbol{X}_s$-t jelöli folytonos s-re, és $\boldsymbol{X}_{s-1}$-et diszkrét s-re. Hasonlóképpen diszkrét s-re az integrál egyszerűen összegzést jelent.

4. Feltételezzük, hogy $\boldsymbol{\theta}_t = \boldsymbol{\theta}$ minden t-re, és Klimko és Nelson (1978) feltételes legkisebb négyzetes (CLS) módszere alapján egy $\widehat{\boldsymbol{\theta}}_T$-vel jelölt becslést adunk a $\boldsymbol{\theta}$ paraméterre.

5. Beírjuk $\boldsymbol{\theta}_t$ helyére $\widehat{\boldsymbol{\theta}}_T$-t az $\boldsymbol{M}_t$ folyamat definíciójában, hogy megkapjuk $\widehat{\boldsymbol{M}}_t^{(T)}$-t.

6. Belátjuk, hogy ha $\boldsymbol{\theta}_t$ állandó t-ben, akkor

$$\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)} := \widehat{\boldsymbol{I}}_T^{-1/2} \widehat{\boldsymbol{M}}_{uT}^{(T)}, \ u \in [0, 1]$$

eloszlásban konvergál egy standard Brown-hídhoz a $[0, 1]$ intervallumon. Itt $\widehat{\boldsymbol{I}}_T$ egy véletlen normáló mátrix, amely a mintából számolható.

7. Ezt felhasználva teszteket definiálunk a $\boldsymbol{\theta}$-ban történő változásra oly módon, hogy $\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)}$ szuprémumát vagy infimumát használjuk tesztstatisztikaként, a változás irányának függvényében.

8. Belátjuk, hogy ha $\boldsymbol{\theta}_t$ a $[0, T]$ intervallumon egyetlen pontban változik, akkor a tesztstatisztikánk $T \to \infty$ mellett sztochasztikusan végtelenhez tart, azaz a tesztünk gyengén konzisztens.

9. Belátjuk, hogy az $\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)}$ folyamat minimum-, illetve maximumhelye jó becslés a $\boldsymbol{\theta}_t$-ben történt változás időpontjára.

Az INAR(p) folyamat

A 2. fejezetben Pap és Szabó (2013) alapján bemutatjuk a p-edrendű egészértékű autoregressziós (INAR(p)) folyamatra elért eredményeket. A folyamat definíciója:

$$X_k = \alpha_1 \circ X_{k-1} + \cdots + \alpha_p \circ X_{k-p} + \varepsilon_k, \ k = 1, 2, \ldots,$$

ahol az ε_k-k független, azonos eloszlású véletlen változók μ várható értékkel, és ha Y nemnegatív egész értékű véletlen változó és $\alpha \in (0, 1)$, akkor $\alpha \circ Y$ jelöli Y db, egymástól és Y-tól is független α várható értékű Bernoulli-eloszlású véletlen változó összegét. A modellt Alzaid és Al-Osh (1987) vezette be $p = 1$-re, majd Du és Li (1991) magasabb p értékekre.

A paramétervektorunk

$$\boldsymbol{\theta} := \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ \mu \end{bmatrix},$$

a stabilitási feltétel pedig

$$\alpha_1 + \ldots + \alpha_p < 1.$$

A (2.4.1) formulában megadott CLS becsléseket és klasszikus martingálelméletet használva egy olyan folyamatot definiálhatunk, amelyről a 2.5.1 Tétel megmutatja, hogy a nullhipotézis mellett egy $p+1$-dimenziós standard Brown-hídhoz tart. Ez alapján a 2.5.1 szakaszban leírtak szerint definiálhatunk változásészlelési eljárásokat.

A teszt gyenge konzisztenciáját a 2.7.1 Tételben látjuk be. A bizonyításban a (2.7.2) felbontást használjuk fel, valamint többször alkalmazzuk a 2.9.1 Lemmát a felbontás elhanyagolható tagjainak becslésére. Ugyanezek az eszközök lesznek alkalmazhatók a változási időpont becslésének aszimptotikus vizsgálatánál is, amelyet a 2.8.1 Tételben teszünk meg.

A Cox–Ingersoll–Ross-folyamat

A 4. fejezet Pap és Szabó (2016) alapján a Cox–Ingersoll–Ross (CIR)-folyamatban történő változásészlelésről szól. A folyamat definíciója:

$$\mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma\sqrt{Y_t}\,\mathrm{d}W_t,\ \ t \in \mathbb{R}_+,$$

ahol $a, b, \sigma > 0$ és $(W_t)_{t\geqslant 0}$ egy standard Wiener-folyamat. A stabilitási feltételünk maga az a-ra és b-re tett megkötés lesz.

A folyamatot először Feller (1951) tanulmányozta, majd Cox et al. (1985) javasolták „short-term” kamatlábmodellként, amelyek közül az egyik legelterjedtebb lett. A folyamat statisztikai vizsgálata így természetesen fontos kérdés volt és sok figyelmet kapott.

A paramétervektorunk a következő lesz:

$$\boldsymbol{\theta} := \begin{bmatrix} a \\ b \end{bmatrix}.$$

Ki fog derülni, hogy a CIR-folyamatra vonatkozó tételek hasonló módon fogalmazhatók meg és bizonyíthatók, mint az INAR(p) folyamat esetén. A (4.1.1) formulában adott becslések nem a szokásosak, de a szerkezetük hasonló az INAR(p) esetben kapott CLS becslésekhez, és a (4.2.1) formulában definiált martingál is hasonló a diszkrét idejű megfelelőjéhez. A nullhipotézis mellett kapott fő eredményünk a 4.2.1 Tétel, melynek bizonyítása ugyanazt az eszközkészletet használja, mint a 2.5.1 Tételé. Az is szembetűnő, hogy a 4.5.1 és 4.6.1 tételek állításai rendkívül hasonlítanak rendre a 2.7.1 és 2.8.1 Tételekre. Ez a bizonyításokban is megjelenik, azonban bizonyos lépések fejlettebb eszközöket kívánnak, mint a diszkrét esetben. Különösen is megemlítjük a 4.7.2 Lemmát, amely a 2.9.1 Lemma helyét veszi át, és amelyet új eredménynek vélünk; továbbá rámutatunk, hogy a 4.7.7 Lemma bizonyítása lényegesen összetettebb, mint a neki megfelelő 2.9.7 Lemmáé.

Paraméterbecslés a Heston-folyamatra

Az 5. fejezetben Barczy et al. (2016) alapján mutatunk be egy CLS módszeren alapuló paraméterbecslést a Heston-modellre, melyet a következő sztochasztikus differenciálegyenlet deifniál:

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma_1\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (\alpha - \beta Y_t)\,\mathrm{d}t + \sigma_2\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big), \end{cases} \qquad t \geqslant 0,$$

ahol $a > 0$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$, és $(W_t, B_t)_{t\geqslant 0}$ egy kétdimenziós standard Wiener-folyamat. Azonnal látható, hogy Y éppen az imént bevezetett Cox–Ingersoll–Ross-folyamat. A stabilitási feltétel itt is, éppúgy, mint a CIR-folyamatra, $b > 0$.

Hamar kiderül, hogy diszkrét megfigyelések alapján CLS becslést adni az (a, b, α, β) paraméterekre igen nehéz, mivel a feltételes várható értékek, következésképpen a parciális deriváltak, összetett módon függnek a paraméterektől. Ezért transzformáljuk a paraméterteret az (5.1.3) függvénnyel, és a transzformált paraméterekre adunk CLS becslést, amihez már lineáris egyenleteket kell megoldanunk. A becsléseket az (5.1.4) formula adja meg. Az 5.2.1 Tételben erős konzisztenciát és aszimptotikus normalitást bizonyítunk ugyanazokkal a martingálelméleti eszközökkel, amelyeket a CIR-folyamatra használtunk (ezeket a 3. fejezetben gyűjtöttük össze). A szükséges számolások azonban sokkal körülményesebbnek bizonyulnak, mint a 4. fejezetben. A CLS becslésekre alkalmazva a transzformáció inverzét, becsléseket kapunk az eredeti paraméterekre. Ezt az (5.4.3) formulában adjuk meg. Ezen becslések erős konzisztenciáját és aszimptotikus normalitását pedig a 5.4.2 Tételben látjuk be, alapozva egyfelől az 5.2.1 Tételre, másfelől az úgynevezett delta-módszerre.

Bibliography

Alzaid, A. A. and Al-Osh, M. A. (1987). First order integer-valued autoregressive (INAR(1)) processes. *J. Time Series Anal.*, 8:261–275.

Baikovicius, J. and Gerencsér, L. (1992). Change-point detection as model selection. *Informatica*, 3(1):3–20.

Barczy, M., Döring, L., Li, Z., and Pap, G. (2014a). Stationarity and ergodicity for an affine two-factor model. *Adv. in Appl. Probab.*, 46(3):878–898.

Barczy, M., Ispány, M., and Pap, G. (2011). Asymptotic behavior of unstable INAR(p) processes. *Stochastic Process. Appl.*, 121(3):583–608.

Barczy, M., Ispány, M., and Pap, G. (2014b). Asymptotic behavior of conditional least squares estimators for unstable integer-valued autoregressive models of order 2. *Scand. J. Stat.*, 41(4):866–892.

Barczy, M. and Pap, G. (2016). Asymptotic properties of maximum-likelihood estimators for Heston models based on continuous time observations. *Statistics*, 50(2):389–417.

Barczy, M., Pap, G., and T. Szabó, T. (2016). Parameter estimation for the subcritical Heston model based on discrete time observations. *Acta Sci. Math.*, 82(1-2):313–338.

Ben Alaya, M. and Kebaier, A. (2013). Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions. *Stoch. Anal. Appl.*, 31(4):552–573.

Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71(5):927–946.

Berkes, I., Gombay, E., Horváth, L., and Kokoszka, P. (2004). Sequential change-point detection in GARCH (p, q) models. *Econometric Theory*, 20(06):1140–1167.

Chung, K. L. (1960). *Markov Chains with Stationary Transition Probabilities.* Springer-Verlag.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-point Analysis.* Wiley.

Csörgő, S. (2010). *Fejezetek a valószínűségelméletből (in Hungarian).* Polygon, Szeged.

Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, 87(3):491–505.

Davis, R. A. and Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika*, 96(3):735–749.

Du, J.-G. and Li, Y. (1991). The integer valued autoregressive (INAR(p)) model. *J. Time Series Anal.*, 12:129–142.

Dudley, R. (2004). *Real Analysis and Probability.* Cambridge University Press.

Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Ann. Probab.*, 14(1):260–270.

Feller, W. (1951). Two singular diffusion problems. *Ann. of Math.*, 54:173–182.

Gerencsér, L. and Prosdocimi, C. (2010). Change detection for Hidden Markov Models. In Edelmayer, A., editor, *19th International symposium on mathematical theory of networks and systems. MTNS, Budapest, 2010.*, pages 1761–1764, Budapest. MTA SZTAKI, ELTE.

Gombay, E. (2001). U-statistics for change under alternatives. *J. Multivariate Anal.*, 78(1):139–158.

Gombay, E. (2008). Change detection in autoregressive time series. *J. Multivariate Anal.*, 99:451–464.

Guo, M. and Härdle, W. K. (2017). Adaptive interest rate modelling. *J. Forecast.*, 36(3):241–256.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343.

Hu, Y. and Long, H. (2007). Parameter estimation for Ornstein–Uhlenbeck processes driven by α-stable Lévy motions. *Commun. Stoch. Anal.*, 1(2):175–192.

Hudecová, Š., Hušková, M., and Meintanis, S. G. (2015). Tests for time series of counts based on the probability-generating function. *Statistics*, 49(2):316–337.

Hurn, A., Lindsay, K., and McClelland, A. (2013). A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions. *J. Econometrics*, 172(1):106–126.

Hušková, M., Prášková, Z., and Steinebach, J. (2007). On the detection of changes in autoregressive time series. I. Asymptotics. *J. Statist. Plann. Inference*, 137(4):1243–1259.

Hyndman, R. J. (n.d.). Time Series Data Library. *http://data.is/TSDLDemo. Accessed on 16 Aug, 2012.*

Ikeda, N. and Watanabe, S. (1989). *Stochastic Differential Equations and Diffusion Processes.* North-Holland.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes.* Springer-Verlag.

Jin, P., Mandrekar, V., Rüdiger, B., and Trabelsi, C. (2013). Positive Harris recurrence of the CIR process and its applications. *Commun. Stoch. Anal.*, 7(3):409–424.

Jin, P., Rüdiger, B., and Trabelsi, C. (2016). Positive Harris recurrence and exponential ergodicity of the basic affine jump-diffusion. *Stoch. Anal. Appl.*, 34(1):75–95.

Kang, J. and Lee, S. (2009). Parameter change test for random coefficient integer-valued autoregressive processes with application to polio data analysis. *J. Time Series Anal.*, 30:239–258.

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus.* Springer-Verlag.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.*, 6:629–642.

Kokoszka, P. and Leipus, R. (1998). Change-point in the mean of dependent observations. *Statist. Probab. Lett.*, 40:385–393.

Kokoszka, P. and Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli*, 6(3):513–539.

Lee, S., Ha, J., Na, O., and Na, S. (2003). The cusum test for parameter change in time series models. *Scand. J. Stat.*, 30(4):781–796.

Lehmann, E. and Romano, J. (2009). *Testing Statistical Hypotheses.* Springer-Verlag.

Li, Z. and Ma, C. (2015). Asymptotic properties of estimators in a stable Cox–Ingersoll–Ross model. *Stochastic Process. Appl.*, 125(8):3196–3233.

Liptser, R. and Shiryaev, A. (2001). *Statistics of Random Processes II. Applications.* Springer-Verlag.

Meyn, S. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability.* Cambridge University Press.

Meyn, S. P. and Tweedie, R. L. (1993). Stability of Markovian processes II: Continuous-time processes and sampled chains. *Adv. in Appl. Probab.*, 25:487–517.

Overbeck, L. (1998). Estimation for continuous branching processes. *Scand. J. Stat.*, 25:111–126.

Overbeck, L. and Rydén, T. (1997). Estimation in the Cox–Ingersoll–Ross model. *Econometric Theory*, 13(3):430–461.

Page, E. (1954). Continuous inspection schemes. *Biometrika*, 41(1-2):100–115.

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3-4):523–527.

Pap, G. and Szabó, T. T. (2013). Change detection in INAR(p) processes against various alternative hypotheses. *Comm. Statist. Theory Methods*, 42(7):1386–1405.

Pap, G. and Szabó, T. T. (2016). Change detection in the Cox–Ingersoll–Ross model. *Statistics & Risk Modeling*, 33(1-2):21–40.

Ploberger, W. and Krämer, W. (1992). The CUSUM test with OLS residuals. *Econometrica*, 60(2):271–285.

Quine, M. P. (1970). The multi-type Galton-Watson process with immigration. *J. Appl. Probab.*, 7(2):411–422.

Schmid, W. and Tzotchev, D. (2004). Statistical surveillance of the parameters of a one-factor Cox–Ingersoll–Ross model. *Sequential Anal.*, 23(3):379–412.

Shiryaev, A. N. (1989). *Probability*. Springer-Verlag.

Silva, I. (2005). *Contributions to the analysis of discrete-valued time series*. PhD thesis, Departamento de Matematica Aplicada, Faculdade de Ciencias da Universidade do Porto.

T. Szabó, T. (2011a). Change point detection for integer-valued autoregressive (INAR(p)) processes. Paper presented at the Hungarian National Conference of Scientific Students' Associations (OTDK).

T. Szabó, T. (2011b). Test statistics for parameter changes in INAR(p) models and a simulation study. *Austrian Journal of Statistics*, 40(4):265–280.

Venkataraman, K. N. (1982). A time series approach to the study of the simple subcritical Galton–Watson process with immigration. *Adv. in Appl. Probab.*, 14(1):1–20.