CHARLES UNIVERSITY
Faculty of mathematics
and physics

HABILITATION THESIS

Šárka Hudecová

Structural tests for time series

Department of Probability and Mathematical Statistics

Mathematics

Probability and Mathematical Statistics

August 29, 2024

I would like to express my deepest thanks to my family for their love and patience. I am grateful to my colleagues at Charles University for their support and to my past teachers and collaborators for everything I have learned from them.

Contents

Preface

This habilitation thesis deals with structural tests for time series, specifically with goodness-of-fit tests and change-point tests. The thesis summarizes the achievements of the following five papers, co-authored by Šárka Hudecová:

[1] Hudecová, Š. (2013): Structural changes in autoregressive models for binary time series. *Journal of Statistical Planning and Inference* **143**(10), 1744–1752.

[2] Hudecová, Š., M. Hušková, and S.G. Meintanis (2015): Tests for time series of counts based on the probability-generating function. *Statistics* **49**(2), 316–337.

[3] Hudecová, Š., M. Hušková, M. and S.G. Meintanis (2017): Tests for structural changes in time series of counts. *Scandinavian Journal of Statistics* **44**(4), 843–865.

[4] Neumeyer, N., M. Omelka, M. and Š. Hudecová (2019): A copula approach for dependence modeling in multivariate nonparametric time series. *Journal of Multivariate Analysis* **171**, 139–162.

[5] Omelka, M., Š. Hudecová, and N. Neumeyer (2020): Maximum pseudo-likelihood estimation based on estimated residuals in copula semiparametric models. *Scandinavian Journal of Statistics* **48**, 1433–1473.

Paper [1] deals with autoregressive models for binary time series and proposes a procedure for detecting a single change in model parameters. Details can be found in Chapter 2.

Articles [2] and [3] introduce structural tests for important classes of count time series models. Goodness–of–fit tests are considered in [2], while an online detection method for an abrupt structural break is proposed and studied in [3]. The content of both papers is summarized in Chapter 3.

Papers [4] and [5] deal with copula modelling of sequences of multivariate observations, and their results are provided in Chapter 4.

The thesis introduces the considered problems and presents the main theoretical results proved in the five articles. The attached papers also contain additional details on computational aspects, practical simulation studies, and real data illustrations, which are not part of this thesis.

The notation in the thesis is unified and, therefore, it need not match the notation used in the attached published versions of papers.

Chapter 1

Introduction

Time series analysis is a statistical technique used to analyze data measured sequentially in time. Such observations are often mutually dependent, so the future behavior of the series can be deduced from the historical data. The aim is to build a suitable model that reasonably describes the underlying structure and enables to estimate quantities of interest, to test hypotheses, and to construct accurate predictions.

Given a specified time series model $\mathcal{M}$ for the observed data, the unknown quantities (often finite dimensional parameters) appearing in the model specification need to be estimated, leading to a fitted model $\widehat{\mathcal{M}}$. This estimated model can then be used for the desired inference tasks. However, an assessment of validity of structural assumptions of model $\mathcal{M}$ is a crucial task because an incorrect model can lead to false conclusions or misleading predictions. Model verification is an objective of *goodness-of-fit testing*, where the aim is to test the null hypothesis

$$\mathcal{H}_0 : \text{the observed data are generated by } \mathcal{M}$$

against a general or a specified alternative $\mathcal{H}_1$.

A model structure can be inappropriate for various reasons. For instance, distributional assumptions might not be feasible and/or the functional dependence on the unknown parameters is not suitable. Different goodness-of-fit tests focus on different types of model violations. Since the world is dynamically evolving, a time series model might be inappropriate due to *time instabilities*. Structural tests, which focus on this type of violation of $\mathcal{H}_0$, belong to the *change point analysis*. Their aim is to evaluate whether the model is stable over the whole observational period, or if there is a change in the data generating process.

This thesis is devoted to structural tests for certain classes of time series.

1.1 Univariate time series models for discrete data

A time series is a sequence of random variables or random vectors measured sequentially in time, typically in equidistant time points. In theory, we typically consider a model for $\{Y_t\}_{t\in\mathbb{Z}}$, where $\mathbb{Z}$ is a set of all integers and, therefore, it is assumed that the process started infinitely in the past and it will continue infinitely in the future. In practice, however, observations from a time window of length T are available, so the data form a finite sequence $\{Y_t\}_{t=1}^T$. In the following we use notation $\{Y_t\} := \{Y_t\}_{t\in\mathbb{Z}}$. We refer to Brockwell and Davis (2006) for basic introduction to time series.

A suitable approach to modeling $\{Y_t\}$ depends on the nature of the individual observations. Consider the univariate situation where $\{Y_t\}$ is a sequence of random variables taking values in $\mathbb{R}$. The traditional modeling approach is based on ARMA or ARMA-GARCH models, see, e.g., Hamilton (1994); Brockwell and Davis (2002); Tsay (2002); Box et al. (2016); Francq and Zakoian (2019). However, this framework is designed for real-valued data. Let

$$\mathcal{F}_t = \sigma\{Y_s,\ s \le t\}$$

stand for the σ-algebra of information known at time t. An autoregressive model AR(p), $p \ge 1$, is defined as

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \varepsilon_t, \tag{1.1}$$

where $\{\varepsilon_t\}$ is a white noise with variance $\sigma^2 > 0$, and $\alpha_i \in \mathbb{R}$, $i = 1,\dots,p$, are model parameters assumed to satisfy certain stationarity and causality conditions. For construction of prediction intervals or for a probabilistic forecast one often has to make a distributional assumption about $\{\varepsilon_t\}$. It is typically assumed that the process $\{\varepsilon_t\}$ is Gaussian, which implies that the conditional distribution of Y_t given $\mathcal{F}_{t-1}$ is normal,

$$Y_t|\mathcal{F}_{t-1} \sim \mathsf{N}\left(\alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}, \sigma^2\right), \tag{1.2}$$

where we use the notation $Z \sim \mathcal{G}$ to specify that a distribution of a random variable Z is $\mathcal{G}$; alternatively, we also write $\mathcal{L}(Z) = \mathcal{G}$.

Various applications require modeling of data $\{Y_t\}$ where Y_t has a discrete distribution on $\mathbb{N}_0$ for each t, see particular examples in Sections 1.1.1 and 1.1.2. When dealing with such *discrete time series*, particularly with those that represent low counts or binary outcomes, then AR models from (1.1) are not suitable, because they neglect the discreteness and they are not able to describe the data adequately. Indeed, since Y_t takes values in $\mathbb{N}_0$, the right hand side of (1.1) would have to be in $\mathbb{N}_0$ for all t, and this would hold only under quite restrictive conditions. Moreover, discrete data often exhibit non-linear relationships and a mean-variance relationship that cannot be captured effectively by ARMA-GARCH models. A solution is to consider special models that account for the discrete nature and specific distributional properties of discrete data. When trying to modify the existing methodologies, one often attempts to recover the basic characteristics of real-valued time series and at the same time borrows useful ideas from the discrete world. In doing so models for discrete time series often mimic the dependence structure of real-valued ARMA (or GARCH) models, while using a formalization that preserves the discreteness.

Assume that $\{Y_t\}$ is a time series of variables taking values in $\mathbb{N}_0$. A natural modification to (1.2) is to replace the normal distribution with a suitable discrete distribution on

$\mathbb{N}_0$ and to assume that

$$Y_t|\mathcal{F}_{t-1} \sim \mathcal{G}(\vartheta_t), \quad \vartheta_t = h(\mathcal{F}_{t-1}, \boldsymbol{\theta}), \tag{1.3}$$

where $\mathcal{G}(\vartheta)$ is a discrete distribution depending on a parameter $\vartheta \in \Omega \subset \mathbb{R}^k$, $k \geq 1$, and h is a link which reflects the dependence of the conditional distribution $\mathcal{G}(\vartheta_t)$ on the past $\mathcal{F}_{t-1}$, typically via some unknown finite dimensional parameter $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$. Several important examples are provided in the next paragraphs.

1.1.1 Binary time series

If Y_t takes only values 0 and 1 for all $t \in \mathbb{Z}$, then the series is formed by *binary* observations. In practice, such data often arise as a sequence of indicators of a presence of a specified event, e.g., precipitation, or whether a stock price was higher or lower compared to the previous day. An example of a binary time series is shown in Figure 1.1, which plots the US quarterly recession data from 1855–2011 obtained from The National Bureau of Economic Research (628 records).

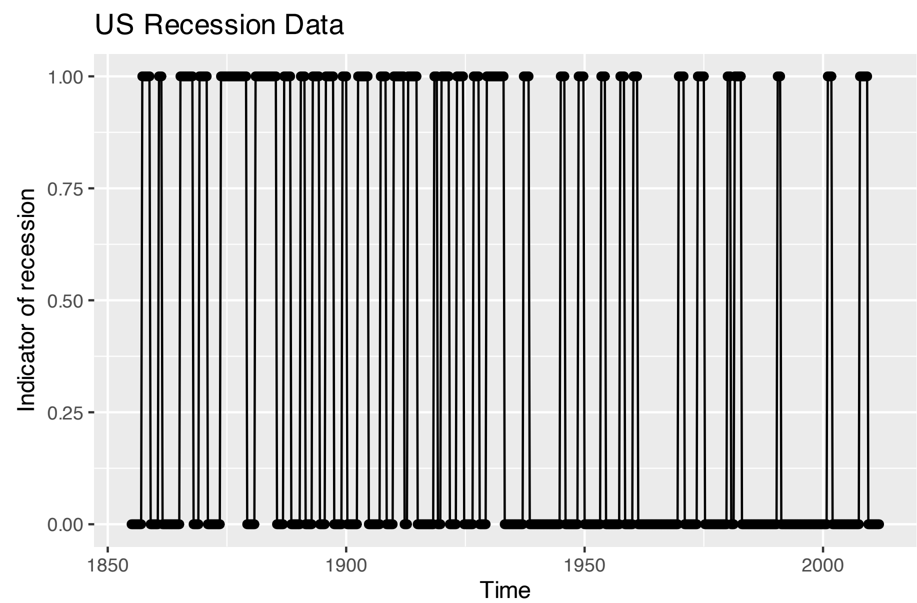

Figure 1.1: An examples of a binary time series: Quarterly indicators of recession in the US. Here, Y_t is coded as 1 if any month in the quarter is being in a recession.

Flexible models for such dependent binary data are based on the principles of generalized linear models (McCullagh and Nelder, 1989; Kedem and Fokianos, 2002). Denote as $\pi_t = \mathsf{P}(Y_t = 1|\mathcal{F}_{t-1})$. Then

$$Y_t|\mathcal{F}_{t-1} \sim \mathsf{B}(\pi_t),$$

where $\mathsf{B}(\pi)$ stands for the Bernoulli distribution with the success probability $\pi \in [0,1]$. A binary autoregressive model of order $p \geq 1$ assumes that

$$\pi_t = h\Big(\beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i}\Big),$$

for a certain specified function $h : \mathbb{R} \to (0,1)$ and some unspecified parameters $\beta_i \in \mathbb{R}$, $i = 0,\ldots,p$. The use of the function h guarantees that the conditional probability π_t lies in $(0,1)$ and the model is well-defined. In the notation of model (1.3), $\mathcal{G}(\vartheta)$ is $\mathsf{B}(\vartheta)$ with $\vartheta \in (0,1) \subset \mathbb{R}$, $\Omega = (0,1)$, and $k = 1$. The unknown parameter of the model is $\boldsymbol{\theta} = (\beta_0,\ldots,\beta_p)^\top$. Such model for binary observations is considered in Chapter 2, where a structural test of a single parametric change point is described.

1.1.2 Count time series

If Y_t has a general discrete distribution on $\mathbb{N}_0$ for all $t \in \mathbb{Z}$ we say that $\{Y_t\}$ is a *count time series*. Such data typically represent counts of specified events occurring in successive time intervals of equal length, as daily numbers of patients, numbers of stock transactions in 5-min intervals, numbers of scored goals in a match of a given team, etc. Examples of such data are provide in Figures 1.2 and 1.3. Figure 1.2 presents a series of monthly polio incidence counts in US from January 1970 to December 1983 (168 observations), (Zeger and Quaqish, 1988; Jung and Tremayne, 2011). Figure 1.3 shows monthly numbers of claims made by injured workers to the British Columbia Workers Compensation Board (burn injuries only) during observation period January 1985 to December 1994, (Freeland, 1998).

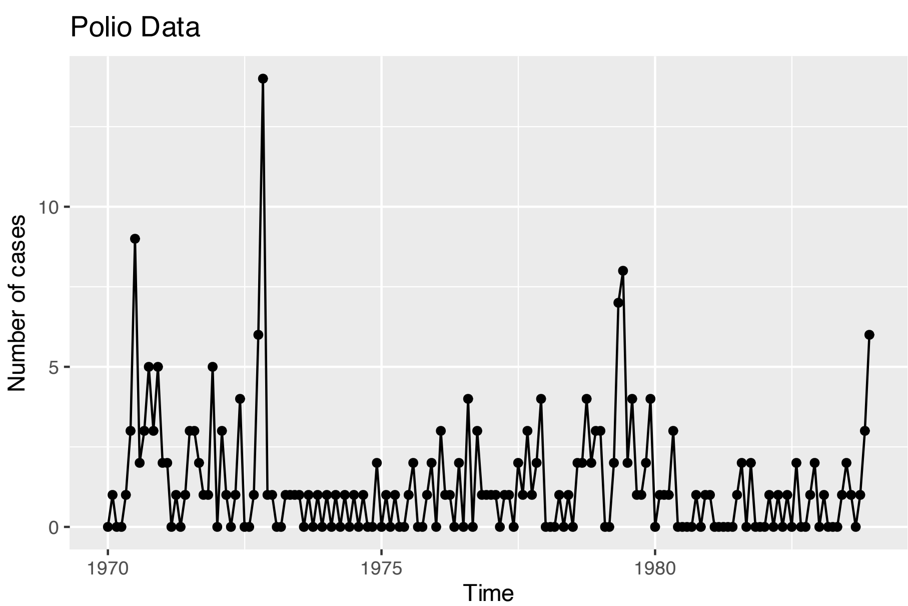

Figure 1.2: Monthly polio incidence counts in the US (Jan 1970 to Dec 1983).

Below we introduce two most important classes of models for such data that are also later used in Chapter 3. We refer to overviews in McKenzie (2003); Davis et al. (2015); Weiss (2018a); Davis et al. (2021) for more details and different approaches.

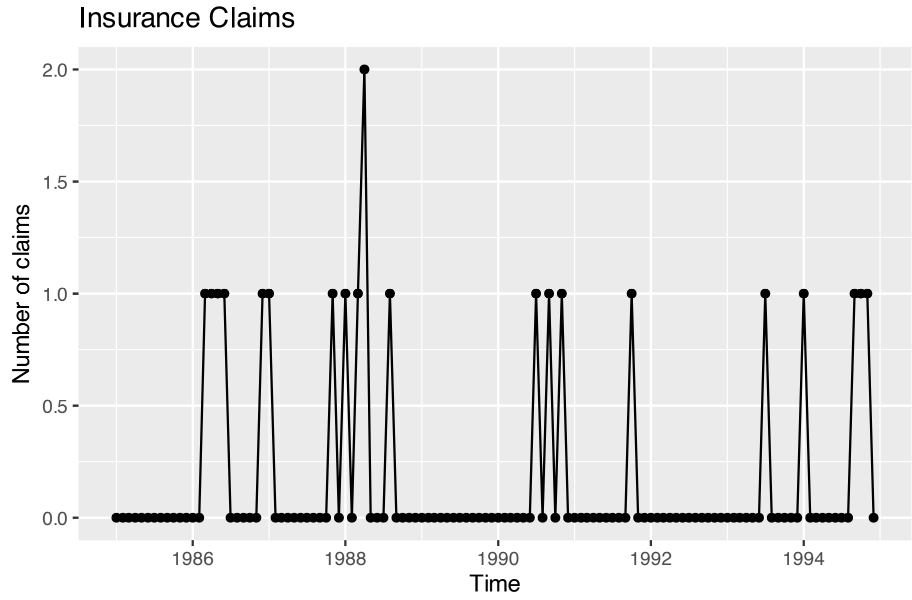

Figure 1.3: Monthly insurance claims (Jan 1985 to Dec 1994).

INGARCH models

For count data, the *Poisson distribution* plays a similar role as the normal distribution in real–valued models because it often serves as a baseline assumption. Hence, one of the basic variants of (1.3) assumes that the conditional distribution $Y_t|\mathcal{F}_{t-1}$ is Poisson with mean λ_t that depends on the past values of the process. Since the mean and variance of the Poisson distribution coincide, that is

$$\lambda_t = \mathsf{E}[Y_t|\mathcal{F}_{t-1}] = \mathsf{var}[Y_t|\mathcal{F}_{t-1}],$$

a model for the conditional mean is also a model for the conditional variance, leading to a *conditionally hereroscedastic model*.

Using similar ideas from generalized linear models combined with the canonical link for Poisson distribution lead to log-linear models, which were considered in various variants by Zeger and Quaqish (1988); Kedem and Fokianos (2002); Davis et al. (2003); Jung et al. (2006); Fokianos and Tjøstheim (2011) among others. An autoregressive model of order $p \geq 1$ can be defined for such data as

$$Y_t|\mathcal{F}_{t-1} \sim \mathsf{Po}(\lambda_t), \quad \log(\lambda_t) = \omega + \sum_{i=1}^{p} \alpha_i Y_{t-i},$$

where $\omega, \alpha_i \in \mathbb{R}$, $i = 1, \dots, p$, are unknown parameters. In notation of (1.3), $k = 1$, $\Omega = (0, \infty)$, $\mathcal{G}(\vartheta) = \mathsf{Po}(\lambda)$ and the unknown parameter is $\boldsymbol{\theta} = (\omega, \alpha_1, \dots, \alpha_p)^\top \in \mathbb{R}^{p+1}$. An identity link together with inclusion of lagged values of λ in the model for λ_t leads to so called INGARCH(p,q) model, $p, q \geq 0$, considered in Ferland et al. (2006), where

$$Y_t|\mathcal{F}_{t-1} \sim \mathsf{Po}(\lambda_t), \quad \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j},$$

where $\omega \in \mathbb{R}$ and $\alpha_i, \beta_j > 0$, $i = 1, \ldots, p$, $j = 1, \ldots, q$. An advantage of the linear model over the log-linear one is that it has analogous second-order properties as the classical ARMA model. A more general variant assumes a general (specified) non-linear dependence of λ_t on its past values and past values of the process (Fokianos and Tjøstheim, 2009, 2012; Fokianos, 2012), which leads to a non-linear INGARCH(p,q) model

$$Y_t|\mathcal{F}_{t-1} \sim \mathcal{G}(\lambda_t), \quad \lambda_t = r(Y_{t-1}, \ldots, Y_{t-p}, \lambda_{t-1}, \ldots, \lambda_{t-q}, \boldsymbol{\eta}),$$

where $\mathcal{G}(\lambda)$ is some distribution on $\mathbb{N}_0$ with mean λ, and the function $r : [0, \infty)^{p+q} \times \Theta$ belongs to some specific parametric family of functions $\mathcal{R} = \{r(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k\}$ for some $k \geq 1$.

Although the Poisson distribution is by far the most popular specification for $\mathcal{G}$, different choices have been considered as well (Zhu, 2011; Christou and Fokianos, 2014; Davis and Liu, 2016), for instance, the conditional negative binomial distribution, the zero-inflated Poisson distribution or a mixture of two Poisson distributions.

Models based on thinning operator

A different class of count time series models is derived as a modification of the classical autoregressive (AR) model for count data. Recall that a causal stationary AR(1) model takes form

$$Y_t = \alpha \cdot Y_{t-1} + \varepsilon_t, \tag{1.4}$$

where $\{\varepsilon_t\}$ is a white noise sequence and $\alpha \in (-1, 1)$ is a parameter. One possibility how to preserve the discreteness of the data is to replace the multiplication in (1.4) with a different operator $\circ$ such that for any random variable Y with values in $\mathbb{N}_0$ the random variable $\alpha \circ Y$ is also discrete with values in $\mathbb{N}_0$.

Integer-valued autoregressive (INAR) models (McKenzie, 1985; Al-Osh and Alzaid, 1987; Alzaid and Al-Osh, 1988; Du and Li, 1991; Scotto et al., 2015) use the Steutel and van Harn's thinning operator $\circ$, which is defined as follows. If Y is a random variable with values in $\mathbb{N}_0$ and $\alpha \in (0, 1)$ then

$$\alpha \circ Y := \sum_{i=1}^{Y} U_i, \tag{1.5}$$

where $\{U_i\}$ are independent and identically distributed (iid) Bernoulli variables with $\alpha = \mathsf{P}(U_i = 1) = 1 - \mathsf{P}(U_i = 0)$, which are independent of Y, with the convention that an empty sum (the case $Y = 0$) equals 0. INAR(1) model then takes the form

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t, \tag{1.6}$$

where $\alpha \in (0, 1)$ is an unknown parameter and $\{\varepsilon_t\}$ is a sequence of iid count random variables with distribution G_ε, independent of the Bernoulli variables involved in the definition of $\circ$. The model can be straightforwardly generalized to INARMA(p,q) models, see Section 3.2.2. If G_ε is Poisson, then the marginal distribution of Y_t is Poisson as well. For this nice property, a model with Poisson G_ε often serves as a baseline model within this class.

Note that model (1.6) can be rewritten as (1.3) via the specification of the conditional distribution of Y_t given $\mathcal{F}_{t-1}$. Namely, $Y_t|\mathcal{F}_{t-1} \sim \mathcal{G}(\vartheta_t)$, where $\mathcal{G}(\vartheta_t)$ is the convolution of G_ε and the binomial distribution $\mathsf{Bi}(Y_{t-1}, \alpha)$.

The two classes of models mentioned above, INGARCH and INARMA, have received enormous attention in the literature. They are known to fit well empirical data in diverse areas of application such as in business, accident prevention, and medicine. However, there exists a number of other models with different structures, see McKenzie (2003). It is, therefore, extremely desirable to have verification tools which would enable a practitioner to validate whether the chosen model is suitable for the data at hand. Chapter 3 presents such structural tests for count time series.

1.2 Multivariate time series models

In various real-life situations, multiple variables are recorded simultaneously in time, resulting in a multivariate time series $\{\boldsymbol{Y}_t\}$ such that $\boldsymbol{Y}_t = (Y_{1t}, \ldots, Y_{kt})^\top$ for some $k > 1$. The components of $\boldsymbol{Y}_t$ are often correlated, so a suitable model for $\{\boldsymbol{Y}_t\}$ has to account for both the mutual dependence among the variables and their time dependence. An example of such multivariate time series is provided in Figure 1.4 for daily exchange rates of US Dollar (USD), British Pound (GBP), and Euro (EUR) to the Czech Koruna (CZK) in 2010–2012 (i.e., $k = 3$) . Each individual series consists of dependent observations, and there are visible trends common to all three series, reflecting their mutual dependence.

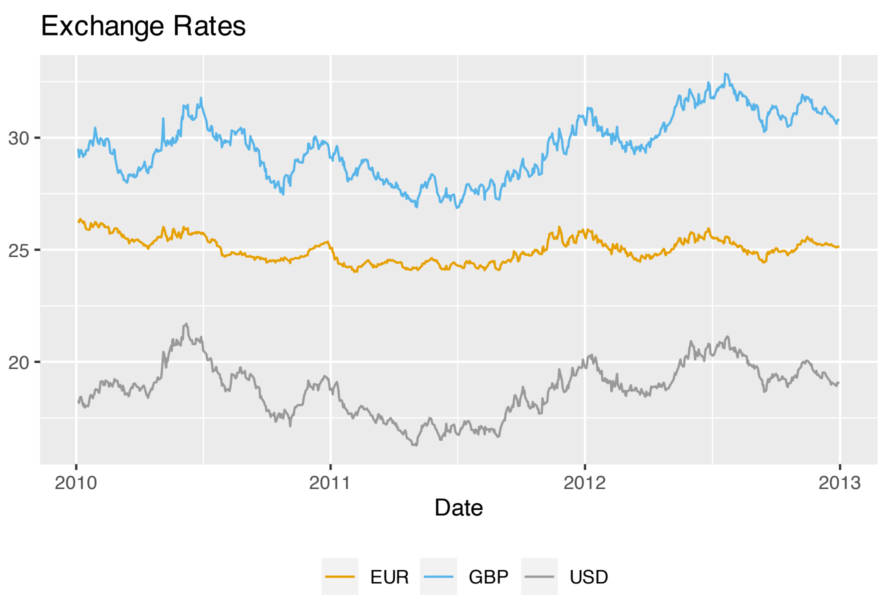

Figure 1.4: Multivariate time series of daily exchange rates of USD/CZK, GBP/CZK, and EUR/CZK in 2010–2012.

The standard approach for modeling stationary multivariate time series is based on multivariate ARMA-GARCH models (Lütkepohl, 2005; Tsay, 2013). These models extend the classical univariate ARMA-GARCH models to the multivariate context, enabling to model interdependencies between multiple univariate time series. However, they are not able to capture certain non-linear relationships among the component variables. Recently, time series models based on *copulas* have proved useful as they allow flexible modeling of complex dependencies and separate the marginal models from the dependence structure.

Modeling of joint distributions using copulas has attained much interest in the literature in the past 20 years (Nelsen, 2006; Embrechts, 2009; Joe, 2014). The approach is based on famous Sklar's theorem (Sklar, 1959) which states that there is a relationship between a cumulative distribution function (cdf) of a random vector $\boldsymbol{X}$ and its marginals, captured by the so called *copula function* or also *copula*. In particular, the cdf H of a random vector $\boldsymbol{X} = (X_1, \dots, X_k)^\top$ can be written as

$$H(x_1,\dots,x_k) = \mathsf{P}(X_1 \le x_1,\dots,X_k \le x_k) = C(F_1(x_1),\dots,F_k(x_k)), \quad \forall (x_1,\dots,x_k) \in \mathbb{R}^k, \tag{1.7}$$

where F_i is the marginal cdf of X_i, $i = 1,\dots,k$, and $C : [0,1]^k \to [0,1]$ is a cdf with uniform marginals, called copula. Moreover, C is uniquely determined on $\times_{i=1}^k \mathrm{Range}(F_i)$, see also (Nelsen, 2006, Chapter 2).

The copulas can be utilized in time series analysis in various ways (Patton, 2009, 2012; Fan and Patton, 2014; Rémillard et al., 2012). Denote as before $\mathcal{F}_t = \sigma\{\boldsymbol{Y}_s,\ s \le t\}$ the information set known at time t. Assume that the conditional distribution of $\boldsymbol{Y}_t$ given $\mathcal{F}_{t-1}$ depends on $\mathcal{F}_{t-1}$ only through a d-dimensional vector $\boldsymbol{X}_t$ for some $d \in \mathbb{N}$, i.e., $\boldsymbol{X}_t$ is $\mathcal{F}_{t-1}$ measurable and $\mathcal{L}(\boldsymbol{Y}_t|\mathcal{F}_{t-1}) = \mathcal{L}(\boldsymbol{Y}_t|\boldsymbol{X}_t)$. For instance, it is often reasonable to consider $\boldsymbol{X}_t = (\boldsymbol{Y}_{t-1}^\top,\dots,\boldsymbol{Y}_{t-p}^\top)^\top$ for some $p \ge 1$. A copula time series model can be written as

$$\boldsymbol{Y}_t = \boldsymbol{m}(\boldsymbol{X}_t) + \Sigma(\boldsymbol{X}_t)\boldsymbol{\varepsilon}_t, \tag{1.8}$$

where

$$\boldsymbol{m}(x) = \Big(m_1(\boldsymbol{x}),\dots,m_k(\boldsymbol{x})\Big)^\top, \quad \Sigma(\boldsymbol{x}) = \mathrm{diag}\Big\{\sigma_1(\boldsymbol{x}),\dots,\sigma_k(\boldsymbol{x})\Big\}$$

for mean functions $m_j : \mathbb{R}^d \to \mathbb{R}$ and volatility functions $\sigma_j : \mathbb{R}^d \to (0,\infty)$, $j = 1,\dots,k$, and $\{\boldsymbol{\varepsilon}_t\}$ are iid random vectors with zero means, unit variances, continuous joint cdf F_ε and marginal cdfs $F_{j\varepsilon}$, $j = 1,\dots,k$. Due to Sklar's theorem, there exists a unique copula C which links the marginal cdfs $F_{j\varepsilon}$, $j = 1,\dots,k$ to the joint F_ε. Hence, model (1.8) can be rewritten using the conditional distribution $\mathcal{L}(\boldsymbol{Y}_t|\mathcal{F}_{t-1})$ as

$$\mathsf{P}(\boldsymbol{Y}_t \le \boldsymbol{y}|\mathcal{F}_{t-1}) = \mathsf{P}(\boldsymbol{Y}_t \le \boldsymbol{y}|\boldsymbol{X}_t) = C\Big(F_{1\varepsilon}(Z_1),\dots,F_{k\varepsilon}(Z_k)\Big), \quad Z_j = \frac{y_j - m_j(\boldsymbol{X}_t)}{\sigma_j(\boldsymbol{X}_t)},\ j = 1,\dots,k,$$

where we use the convention that an inequality $\boldsymbol{x} \le \boldsymbol{y}$ for two vectors $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^k$ is meant componentwise, that is $\boldsymbol{x} \le \boldsymbol{y}$ if and only if $x_i \le y_i$ for all $i = 1,\dots,k$, where $\boldsymbol{x} = (x_1,\dots,x_k)^\top$ and $\boldsymbol{y} = (y_1,\dots,y_k)^\top$. Thus, the copula C reflects the conditional dependence structure of the individual k series after removing the influence of $\boldsymbol{X}_t$.

Hence, the following three components of the model (1.8) need to be estimated: (i) the individual mean and volatility functions m_j and σ_j, $j = 1,\dots,k$, (ii) the marginal cdfs of the innovations $F_{j\varepsilon}$, $j = 1,\dots,k$, and (iii) the copula function. The marginal cdfs in (ii) are typically estimated nonparametrically, see Chapter 4, while it is often suitable to specify a parametric model for the copula in (iii) as

$$C \in \mathcal{C} = \{C_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}, \tag{1.9}$$

where $\mathcal{C}$ can be, for example, a family of Gaussian or Clayton copulas, see (Joe, 2014, Chapter 4).

Different copula models capture different types of dependence structures, as illustrated by Figure 1.5 which plots $n = 500$ independent copies of a random vector $(Z_1, Z_2)^\top$ with a bivariate distribution with the standard normal marginals $\mathcal{N}(0,1)$ and two different copulas C. In both cases, the Kendall's tau is $\tau = 0.5$. The left panel corresponds to C being a Clayton copula (with parameter $\theta = 2$), while the right panel provides samples from the Gaussian copula (with $\theta \doteq 0.7$). One can observe different tail dependencies (extreme co-movements) for the two different copulas.

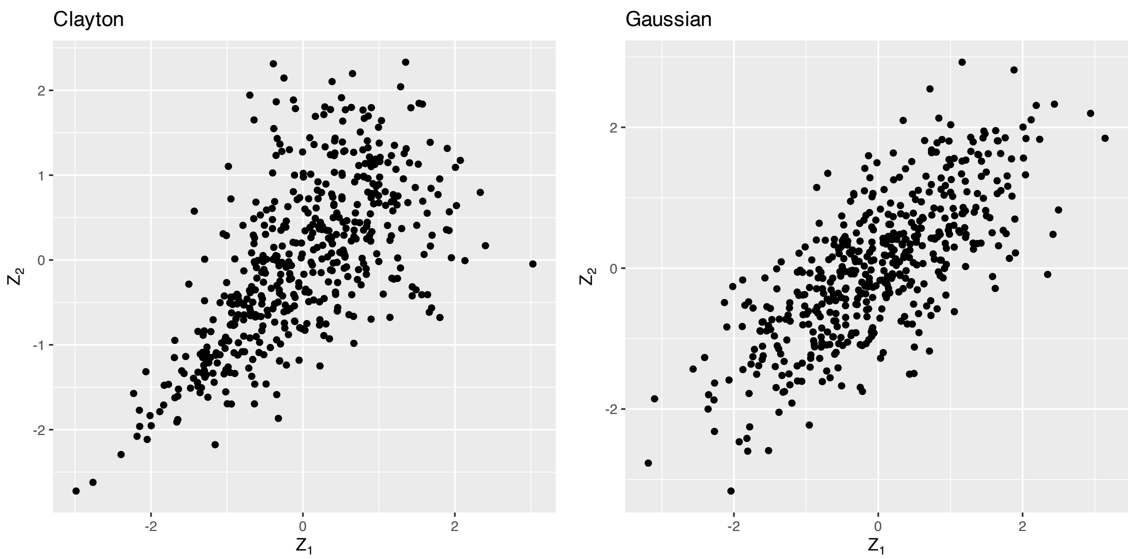

Figure 1.5: Illustration of different dependence structures for different copula families: $n = 500$ independent random vector sampled from a distribution with the standard normal marginals and a copula C: Clayton copula (left panel) and Gaussian copula (right panel). The Kendall's tau is $\tau = 0.5$ in both cases.

An inappropriate copula can lead to inaccurate description of the relationships among the variables, especially their tail dependencies that are crucial, e.g., in risk management and financial applications. Selecting appropriate copula model is crucial because it directly impacts the quality and reliability of the resulting multivariate time series model (1.8). A verification of the assumption (1.9) via a structural goodness-of-fit test is, therefore, desirable in all applications. This setup is considered in Chapter 4.

1.3 Goodness-of-fit testing

Goodness-of-fit testing is a statistical method used to assess how well the observe data $Y_1, \ldots, Y_T$ fit to a specified parametric model $\mathcal{M}$. A difficulty in testing such a hypothesis is that the class of alternatives is typically enormously large (Lehmann et al., 2005, Chapter 14).

The goodness-of-fit tests for iid univariate continuous data typically employ the cumulative distribution function (cdf) and its empirical counterpart. Assume a parametric model $\mathcal{M}$ which specifies cdf $F_{\boldsymbol{\theta}}$ for Y_t for some unknown parameter $\boldsymbol{\theta} \in \Theta$. Let $\widehat{F}_T$ be the empirical cdf and $\widehat{\boldsymbol{\theta}}$ be a consistent estimator of $\boldsymbol{\theta}$ constructed from $Y_1, \dots, Y_T$ under the assumption that $\mathcal{M}$ holds. Then a typical goodness-of-fit test statistic is based on a distance $d(\widehat{F}_T, F_{\widehat{\boldsymbol{\theta}}})$ for a suitable metric d. Most popular choices are

$$S_{T,1} = \sup_{x \in \mathbb{R}} \left| \widehat{F}_T(x) - F_{\widehat{\boldsymbol{\theta}}}(x) \right| \tag{1.10}$$

leading to the Kolmogorov-Smirnov test statistic, and

$$S_{T,2} = \int_{-\infty}^{\infty} \left(\widehat{F}_T(x) - F_{\widehat{\boldsymbol{\theta}}}(x) \right)^2 w(x) \mathrm{d}x, \tag{1.11}$$

where $w : \mathbb{R} \to (0, \infty)$ is an integrable weight function, leading to the Cramér–von Mises test statistic. The model $\mathcal{M}$ is rejected if the chosen test statistic S_T (being either $S_{T,1}$ or S_{T_2} or some other distance) exceeds some threshold, let say c. This bound $c = c_\alpha$ has to be chosen such that the test keeps a prescribed (asymptotic) level $\alpha \in (0,1)$. It can be determined as the corresponding quantile of the asymptotic distribution of S_T.

When dealing with discrete distributions, the use of a cumulative distribution function is often inconvenient due to its non-continuous nature. See, for instance, Gürtler and Henze (2000) where the application of the Cramér–von Mises test is shown to be non–trivial for the Poisson distribution. A modification of the Pearson's χ^2 test can be used for discrete data (Lehmann et al., 2005, Chapter 14.3), but various practical issues arise. A different approach is to replace the cdf by a different characteristic. An interesting possibility is to use the *probability generating function* (PGF), introduced in Section 3.1, that has been successfully used for testing goodness-of-fit for iid data in various situations, e.g., in Baringhaus and Henze (1992); Nakamura and Pérez-Abreu (1993); Kocherlakota and Kocherlakota (1986). In the iid setup, reasonable test statistics can be obtained if the cdfs are replaced with PGFs in (1.10) and (1.11). We follow this approach in Section 3.3 and adapt it to the time series setup. The crucial task is to derive the formula for the critical value c_α.

Verification of copula models was considered by various authors, see Fermanian (2005), Genest and Rémillard (2008), Genest and Rémillard (2006), Omelka et al. (2009) for the iid case, Rémillard (2017) for time series, and Fermanian (2013) and Genest et al. (2009) for some overviews. Since every copula is, in fact, a multivariate cdf, one can consider analogous versions of the Kolmogorov-Smirnov or Cramér von Mises test statistics. However, in the considered time series model (1.8), C is not a cdf of the *observed* data, but rather a cdf of *unobserved* innovations that depend on observations via an unknown transformation. Namely, C is the cdf of $\left(F_{1\varepsilon}(\varepsilon_{1t}), \dots, F_{k\varepsilon}(\varepsilon_{kt})\right)^\top$, where $F_{j\varepsilon}$, $j = 1, \dots, k$, are unknown and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{kt})^\top$ are not directly observable. An empirical estimator $\widehat{C}_T$ can be constructed from the model residuals, but a crucial task is to show that the estimation of the marginal means and volatilities together with the estimation of the marginal cdfs $F_{j\varepsilon}$ do not affect the asymptotic behavior of the corresponding goodness-of-fit test statistics. This problem is considered in Chapter 4.

1.4 Change point detection

Stability of a time series model over a longer time period is always questionable. It is often assumed that the system under consideration may change at some unknown time moments, so called *change points*, while it remains stable between these points. The main objective is then to decide whether a change has occurred, and if this is the case, to estimate the change points. Detecting change points is crucial in various fields, including signal processing, quality control, finance, and environmental monitoring.

There is a broad statistical literature related to the change point topic with (Csörgő and Horváth, 1997) being one of the basic references. In the time series setting, various results have been derived for ARMA models (e.g., Horváth, 1993; Davis et al., 1995; Hušková et al., 2007) and GARCH models (e.g., Kokoszka and Teyssière, 2002; Berkes et al., 2004). Ling (2007) deals with the detection of changes in general time series models (which include ARMA and GARCH models as special cases) under the near epoch dependence assumption. Aue and Horváth (2013) provide an extensive overview of the recent work in this field. More specific references to the literature on the change point problem within count time series are provided in the introduction to Chapter 3.

We distinguish between *offline* and *online* (also *sequential*) monitoring, two different approaches to detecting changes over time.

In *offline change detection*, the entire dataset is available before the analysis begins. Assume that data $Y_1, \ldots, Y_T$ are available and we intend to test that a change point has occurred within the observation period $\{1, \ldots, T\}$. The basic situation of a single change aims to test

$$\mathcal{H}_0 : Y_t \text{ follows model } \mathcal{M} \text{ for all } t = 1, \ldots, T,$$

against

$$\mathcal{H}_1 : Y_t \text{ follows } \mathcal{M}_1 \text{ for } 1 \le t \le t_0 \text{ and } \mathcal{M}_2 \text{ for } t_0 < t \le T,$$

where $t_0 \in \{2, \ldots, T-1\}$ is an unknown change point and $\mathcal{M}_1 \neq \mathcal{M}_2$, for instance, due to a parametric or distributional change. The test is based on a test statistic $S_T = S_T(Y_1, \ldots, Y_T)$ and $\mathcal{H}_0$ is rejected if $S_T > c_\alpha$ where c_α is the critical value which ensures that the test has asymptotic level α, i.e.,

$$\lim_{T \to \infty} \mathsf{P}_{\mathcal{H}_0}(S_T > c_\alpha) = \alpha$$

for a specified $\alpha \in (0, 1)$, typically $\alpha = 0.05$ in practice. Such test of $\mathcal{H}_0$ often naturally induces an estimator of the unknown change point t_0. This kind of change point problem is considered in Chapter 2 for binary time series.

In the *sequential (online) testing* the observations arrive sequentially and the test is implemented anew with each new arriving observation. It is assumed that for some $T > 0$, there exists a *training sample* (historical data) $Y_1, \ldots, Y_T$ that involves no change. The monitoring period begins with time $t = T + 1$ and continues till time $T(m + 1)$ for some

$m \in \mathbb{N} \cup \{\infty\}$. If $m < \infty$, the procedure is called *close–end*, while for $m = \infty$ it is referred to as an *open–end* procedure. Hence, the situation is as follows:

$$\underbrace{Y_1, \dots, Y_T}_{\text{training data}} \quad \underbrace{\overset{}{\underset{\uparrow}{Y_{T+1}}}, \underset{\uparrow}{Y_{T+2}}, \dots, \underset{\uparrow}{Y_{T(m+1)}}}_{\text{monitoring period of length } mT}$$

Consider

$$\mathcal{H}_0: \; Y_t \text{ follows model } \mathcal{M} \text{ for all } t \in \mathbb{Z}$$

against

$$\mathcal{H}_1: \; Y_t \text{ follows model } \mathcal{M}_1 \text{ for all } t \le T + t_0 \text{ and } \mathcal{M}_2 \text{ for } t > T + t_0,$$

where $t_0 > 1$ is an unknown change point and $\mathcal{M}_1 \neq \mathcal{M}_2$. The null hypothesis is rejected if the value of a suitable detector statistic exceeds an appropriately chosen bound *for the first time*. If the detector statistic at time $T + t$ is below the threshold, we continue monitoring and recompute the detector at time $T + t + 1$ etc.

A detector statistic is typically of the form

$$D_{T,t,\gamma} = \frac{\widehat{\Delta}_{T,t}}{q_\gamma^2\left(\frac{t}{T}\right)}, \quad t \ge 1, \tag{1.12}$$

where $\widehat{\Delta}_{T,t}$ is a suitable function of the observations $Y_1, \dots, Y_{T+t}$ (sensitive to violations of $\mathcal{H}_0$) and where q_γ is a weight function,

$$q_\gamma(s) = (1+s)\left(\frac{s}{s+1}\right)^\gamma, \; \gamma \in [0, 1/2). \tag{1.13}$$

The parameter γ in (1.13) is a tuning parameter, see Aue and Horváth (2004) and Zeileis (2005). Briefly, a value of γ close to 0 is appropriate for a late change while if an early change is expected, γ should be taken close to 1/2.

The corresponding stopping rule is specified as

$$\tau(T; m) = \tau(T) = \begin{cases} \inf\left\{1 \le t \le mT : D_{T,t,\gamma} > c_{\alpha,m}\right\}, \\ +\infty, \text{ if } D_{T,t,\gamma} \le c_{\alpha,m} \text{ for all } 1 < t \le mT, \end{cases} \tag{1.14}$$

where $c_{\alpha,m}$ is a constant that guarantees that the test has asymptotic size equal to $\alpha \in (0,1)$. That is

$$\lim_{T\to\infty} \mathsf{P}_{\mathcal{H}_0}\Big(\tau(T) < \infty\Big) = \lim_{T\to\infty} \mathsf{P}_{\mathcal{H}_0}\left(\max_{1\le t\le mT} D_{T,t,\gamma} > c_{\alpha,m}\right) = \alpha. \tag{1.15}$$

Under the alternative $\mathcal{H}_1$ we require

$$\lim_{T\to\infty} \mathsf{P}_{\mathcal{H}_1}\Big(\tau(T) < \infty\Big) = 1. \tag{1.16}$$

The crucial task is to determine the critical value $c_{\alpha,m}$ for $\alpha \in (0,1)$, $m > 0$ and for a given detector test statistic $D_{T,t,\gamma}$ so that (1.15) holds. If $\mathcal{H}_0$ is rejected, then $\tau(T)$, the time when the detector $D_{T,t,\gamma}$ crosses the threshold $c_{\alpha,m}$ for the first time, is taken as an estimator of the unknown time of change t_0.

The described sequential change point detection is illustrated graphically in Figure 1.6. In this example, we set $T = 50$ and $m = 4$, so we have a historical data of length 50 and the monitoring is conducted sequentially on observations Y_{T+t} for $t = 1,\ldots,200 = mT$. In each step t, a detector statistic $D_{T,t,\gamma}$ is computed and compared with the red boundary $c_{\alpha,m}$, which is determined such that (1.15) holds. The two plots show 20 trajectories of $\{D_{T,t,\gamma}\}_{t=1}^{mT}$ simulated under $\mathcal{H}_0$ (left panel) and under $\mathcal{H}_1$ with one change point at $t_0 = 50$ (right panel). Only 1 out of 20 trajectories simulated under $\mathcal{H}_0$ exceeds the boundary $c_{\alpha,m}$ (this realization is highlighted with a dark gray color), which illustrates the validity of (1.15). Only 3 out of 20 trajectories simulated under $\mathcal{H}_1$ do not exceed $c_{\alpha,m}$ (stressed by a dark gray color), which is in agreement with the requirement (1.16). The time point $\tau(T)$, when the detector $D_{T,t,\gamma}$ crosses the red boundary for the first time, is the estimator of the unknown t_0. In this example, its values range between 63 and 142.

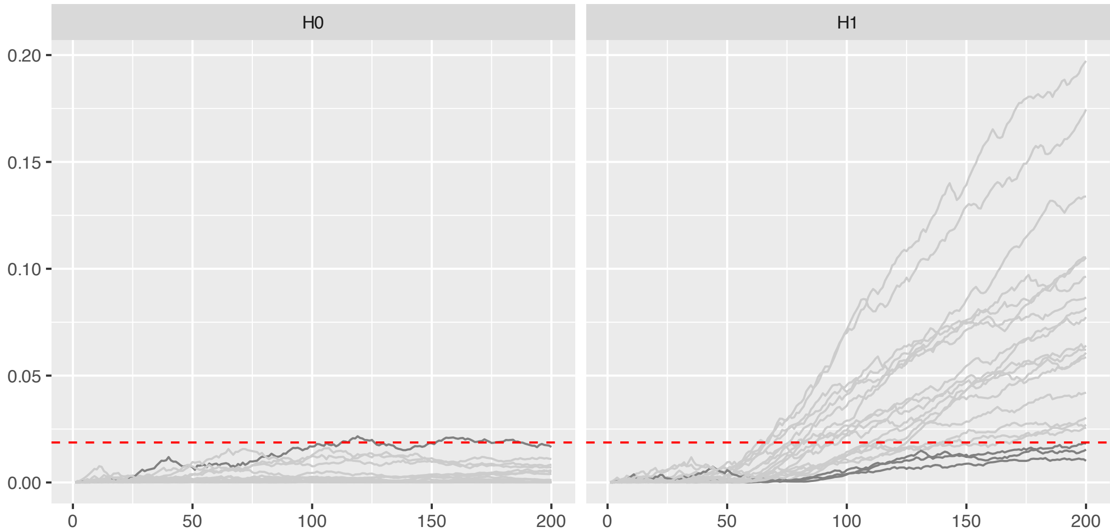

Figure 1.6: An illustration of a sequential detection with $T = 50$ and $m = 4$: 20 trajectories of $\{D_{T,t,\gamma}\}_{t=1}^{mT}$ simulated under $\mathcal{H}_0$ (left panel) and under $\mathcal{H}_1$ with a change point at $t_0 = 50$ (right panel).

Procedures satisfying (1.15) and (1.16) were introduced and studied by Chu et al. (1996) in the context of simple parameter change, and were further extended and modified in, e.g., Leisch et al. (2000), Horváth et al. (2004), Zeileis (2005), Aue and Horváth (2004), and Aue et al. (2006). We consider this type of detection in Section 3.4.

Chapter 2

Instabilities in Binary Time Series

Binary time series play an important role in many fields of application. They are typically observed when one is concerned with an occurrence of an event in a time period. For example, daily occurrences of precipitation can be modeled as binary time series (see Wilks and Wilby, 1999). In a financial area, one might be interested in recession indicators (see Kauppi and Saikkonen, 2008; Startz, 2008), or in a series of direction-of-change of stock returns.

Detection of changes in the success probability of independent binary variables was studied by Pettitt (1980) using a CUSUM type test statistic. Changes in the success probability of independent binomial variables were further considered, e.g., by Worsley (1983); Horváth (1989); Ma (1997); Serbinowska (1996). In this chapter we deal with the change point problem within the framework of dependent binary variables.

Section 2.1 introduces an autoregressive model for binary time series which serves as the baseline model for our study. The procedure for a detection of a change in the model parameters is introduced in Section 2.2. The derived test statistic is a maximum of normalized sums of the estimated residuals, but the normalization is slightly more complicated compared to the standard CUSUM type statistic. Our procedure is closely related to the problem of detection of changes in generalized linear models studied by Antoch et al. (2004). Due to the form of the test statistic, the test is sensitive to any change in the model which leads to a change in the unconditional success probability.

This chapter introduces theoretical results from Hudecová (2013b).

2.1 Autoregressive models for binary time series

Let $\{Y_t\}$ be a binary (0-1 valued) time series of interest, and let $\mathcal{F}_{t-1} = \sigma\{Y_s, s \le t-1\}$ be the σ-field generated by the past $\{Y_s, s \le t-1\}$. Assume that the conditional distribution of Y_t given $\mathcal{F}_{t-1}$ is binary $\mathsf{B}(\pi_t)$ such that the success probability π_t depends on p previous values of the series via the model

$$g(\pi_t) = \beta_0 + \beta_1 Y_{t-1} + \ldots \beta_p Y_{t-p}, \tag{2.1}$$

where g is a suitable link function (logit, probit) and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ is a vector of unknown parameters. Equivalently,

$$\pi_t = g^{-1}\left(\beta_0 + \beta_1 Y_{t-1} + \ldots \beta_p Y_{t-p}\right). \tag{2.2}$$

This model is referred to as *a binary autoregressive model* (BAR), (e.g., Wang and Li, 2011), or *a binary dynamic response model*, (Kauppi and Saikkonen, 2008; de Jong and Woutersen, 2011). It is shown by Wang and Li (2011) that there always exists a stationary solution to (2.1).

The model (2.1) can be further extended by including explanatory variables or lagged values of π_t, (Kauppi and Saikkonen, 2008; Startz, 2008). All these models then belong to a wide class of models, called *time series following generalized linear models*, (Kedem and Fokianos, 2002).

Let $Y_1, Y_2, \ldots, Y_n$ be the observed data. The vector of parameters $\boldsymbol{\beta}$ from (2.1) can be estimated by the conditional maximum likelihood method, so that the estimator $\widehat{\boldsymbol{\beta}}$ is the maximizer of the conditional likelihood function

$$L(\boldsymbol{\beta}) = \prod_{t=p+1}^{n} [\pi_t(\boldsymbol{\beta})]^{Y_t} [1 - \pi_t(\boldsymbol{\beta})]^{1-Y_t},$$

where $\pi_t(\boldsymbol{\beta}) = \pi_t$ is given by (2.2). Alternatively, the estimator $\widehat{\boldsymbol{\beta}}$ can be computed as

$$\widehat{\boldsymbol{\beta}} = \operatorname{argmax}_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \prod_{t=1}^{n} [\pi_t(\boldsymbol{\beta})]^{Y_t} [1 - \pi_t(\boldsymbol{\beta})]^{1-Y_t}, \tag{2.3}$$

for some initial values $Y_{-p+1}, \ldots, Y_0$.

The estimator $\widehat{\boldsymbol{\beta}}$ is almost surely unique for all sufficiently large sample size n, consistent, and asymptotically normal under some regularity conditions (Kedem and Fokianos, 2002, Chap. 1). This estimation approach is convenient in applications, because software tools available for generalized linear models can be directly used. Testing hypotheses about $\boldsymbol{\beta}$ is based on the conditional maximum likelihood as well. The common tests are based on likelihood ratio statistic, the Wald statistic, and the score statistic. The limiting distribution of all these test statistics under the null hypothesis is a χ^2, similarly as in the classical maximum likelihood inference.

In the following, we consider the logit link function $g(x) = \operatorname{logit}(x) = \log[x/(1-x)]$ in model (2.1), because this is the canonical link for a binary regression, and this fact simplifies some of the formulas. Results for a different link function $g : (0,1) \to \mathbb{R}$ could be derived in the same way, provided that g satisfies some standard regularity conditions, (Kedem and Fokianos, 2002, Chap. 3).

2.2 Testing for a change

In order to simplify the notation, assume that $Y_{-p+1}, \ldots, Y_1, Y_2, \ldots, Y_n$ are data generated by model (2.1). This means that we have n realizations of $(Y_t, Y_{t-1}, \ldots, Y_{t-p})^\top$. We would like to decide whether a change has appeared in the data generating process. Hence, we introduce the model

$$\text{logit}(\pi_t) = \begin{cases} \beta_0 + \sum_{j=1}^{p} \beta_j Y_{t-j}, & t = 1, \ldots, m \\ \beta_0^* + \sum_{j=1}^{p} \beta_j^* Y_{t-j}, & t = m+1, \ldots, n, \end{cases} \tag{2.4}$$

where $\boldsymbol{\beta} \neq \boldsymbol{\beta}^*$. Model (2.4) describes the situation where the first m observations follow the model (2.1) with the parameters $\boldsymbol{\beta}$, and the remaining $n-m$ observations follow the model (2.1) with the parameters $\boldsymbol{\beta}^*$. The main objective is to test whether a change has occurred or not, i.e., to test

$$\mathcal{H}_0 : m = n \quad \text{against } \mathcal{H}_1 : m < n. \tag{2.5}$$

We derive the test statistic for $\mathcal{H}_0$ against a simplified alternative of a change in the intercept only, that is for the case where (2.4) holds with $\beta_0 \neq \beta_0^*$ and $\beta_j = \beta_j^*$ for $j = 1, \ldots, p$.

Assume first that the change point is known, let say $m = k$ for some $k < n$. Consider the model

$$\text{logit}(\pi_t) = \begin{cases} \beta_0 + \delta_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} & t = 1, \ldots, k \\ \beta_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} & t = k+1, \ldots, n, \end{cases}$$

and derive the score test statistic, denoted as $c_n^{(k)}$, for the test of $\mathcal{H}_0^* : \delta_0 = 0$ against $\mathcal{H}_1^* : \delta_0 \neq 0$ in this model. Let $\widehat{\boldsymbol{\beta}}$ be the estimated vector of model parameters computed under the null hypothesis of no change as (2.3) and $\widehat{\pi}_t$ be the corresponding estimated conditional mean of Y_t given $\mathcal{F}_{t-1}$ computed from (2.2) with $\boldsymbol{\beta}$ replaced with $\widehat{\boldsymbol{\beta}}$. Define

$$\widehat{\sigma}_t^2 = \widehat{\pi}_t(1 - \widehat{\pi}_t).$$

After some computation, it follows that

$$c_n^{(k)} = \frac{\left[\sum_{t=1}^{k} (Y_t - \widehat{\pi}_t)\right]^2}{V_k}, \tag{2.6}$$

where

$$V_k = \sum_{t=1}^{k} \widehat{\sigma}_t^2 - \left[\sum_{t=1}^{k} \widehat{\sigma}_t^2 \boldsymbol{Z}_{t-1}\right]^\top \left(\sum_{t=1}^{n} \widehat{\sigma}_t^2 \boldsymbol{Z}_{t-1} \boldsymbol{Z}_{t-1}^\top\right)^{-1} \left[\sum_{t=1}^{k} \widehat{\sigma}_t^2 \boldsymbol{Z}_{t-1}\right], \tag{2.7}$$

and $\boldsymbol{Z}_{t-1} = (1, Y_{t-1}, \ldots, Y_{t-p})^\top$.

If the (possible) change point m is known then the test of $\mathcal{H}_0$ of no change in (2.4) can be performed easily using the score statistic $c_n^{(k)}$ and its asymptotic χ_1^2 distribution. In practice, the time point of change is usually unknown and, thus, a natural idea is to base the test of $\mathcal{H}_0$ on the maximum of $c_n^{(k)}$ over all possible k. Set

$$\widehat{S}_k = \sum_{t=1}^{k} (Y_t - \widehat{\pi}_t), \quad k = 1, \ldots, n,$$

the cumulative sums (CUSUM) of residuals $Y_t - \widehat{\pi}_t$. Let k_0 be such that V_k is well defined and positive for all $k_0 \leq k \leq n - k_0$. Define the maximal score test statistic

$$T_n = \max_{k_0 \leq k \leq n-k_0} \sqrt{c_n^{(k)}} = \max_{k_0 \leq k \leq n-k_0} \frac{|\widehat{S}_k|}{\sqrt{V_k}}. \tag{2.8}$$

Furthermore, consider also the following variants with different normalization:

$$U_n = \max_{k_0 \le k \le n-k_0} \left\{ \sqrt{\frac{\sum_{t=1}^{n} \widehat{\sigma}_t^2}{\sum_{t=1}^{k} \widehat{\sigma}_t^2 \sum_{t=k+1}^{n} \widehat{\sigma}_t^2}} |\widehat{S}_k| \right\}, \tag{2.9}$$

$$W_n = \max_{k_0 \le k \le n-k_0} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{|\widehat{S}_k|}{\sqrt{\frac{1}{n}\sum_{t=1}^{n} \widehat{\sigma}_t^2}} \right\}. \tag{2.10}$$

All statistics T_n, U_n, and W_n are maxima of normalized cumulative sums of residuals $Y_t - \widehat{\pi}_t$, but they differ in the normalization. Clearly, the power of the corresponding test is the largest against the alternative of a change in β_0. However, the test is sensitive to any change in $\boldsymbol{\beta}$ which leads to a change in the unconditional success probability $\mathsf{E}\pi_t$, and, thus, it can be applied even in the general case (2.4), as illustrated in a simulation study in Hudecová (2013b). Moreover, in many applications, it is common that any change in $\boldsymbol{\beta}$ is accompanied with a change in β_0.

In order to derive the asymptotic distribution of the test statistics T_n, U_n, and W_n under the null hypothesis $\mathcal{H}_0$, we need to consider the following assumptions:

(A1) The true parameter $\boldsymbol{\beta}$ lies in an open subset of $\mathbb{R}^{p+1}$.

(A2) The series $\{Y_t\}$ is strictly stationary.

Theorem 2.1. *Let assumptions (A1)–(A2) hold. Then under $\mathcal{H}_0 : m = n$ it holds that*

$$\mathsf{P}\left(T_n < \sqrt{2\log\log n} + \frac{\log\log\log n}{2\sqrt{2\log\log n}} + \frac{t - \frac{1}{2}\log\pi}{\sqrt{2\log\log n}} \right) \to \exp\Big\{-2\exp(-t)\Big\} \tag{2.11}$$

as $n \to \infty$, $t \in \mathbb{R}$, and the test statistics U_n and W_n have the same asymptotic distribution as T_n.

The test of the null hypothesis of no change $\mathcal{H}_0 : m = n$ can be based on a comparison of the chosen test statistic (either T_n, U_n or W_n) with the asymptotic critical value c_α, which can be easily computed from the limiting distribution (2.11). In particular, denote

$$a_n = \sqrt{2\log\log n} + \frac{\log\log\log n}{2\sqrt{2\log\log n}}, \qquad b_n = \sqrt{2\log\log n},$$

and $G(t) = \exp\{-2\mathrm{e}^{-t}\}$ for $t \ge 0$. For $\alpha \in (0,1)$ set

$$t_\alpha = G^{-1}(\alpha) = -\log\left[-\frac{1}{2}\log(1-\alpha)\right]$$

and

$$c_\alpha = a_n + \frac{t_\alpha - \frac{1}{2}\log\pi}{b_n}.$$

Then it follows from (2.11) that

$$\lim_{n\to\infty} \mathsf{P}(T_n > c_\alpha) = \alpha$$

and the same holds if T_n is replaced with U_n or W_n. Hence, we reject $\mathcal{H}_0$ if the value of the chosen test statistic exceeds c_α. If the test statistic T_n is chosen and $\mathcal{H}_0$ is rejected, then

$$\widehat{m} = \operatorname{argmax}_{k_0 \le k \le n-k_0} \frac{|\widehat{S}_k|}{\sqrt{V_k}}$$

is an estimator of the unknown change point m. Similarly for tests based on statistics U_n and W_n.

Remark 2.1. The test statistics in (2.8)–(2.10) are defined as normalized sums of estimated residuals $Y_t - \widehat{\pi}_t$. They are, therefore, sensitive to a change in the unconditional success probability $\mathsf{E}Y_t = \mathsf{E}\pi_t$. If $\boldsymbol{\beta} \neq \boldsymbol{\beta}^*$, but the corresponding success probabilities satisfy $\mathsf{E}\pi_t = \mathsf{E}\pi_t^*$ then the test is not able to detect the change. This problem of test statistics based on estimated residuals is well known in the linear regression, (Hušková and Koubková, 2005; Horváth et al., 2004). The same situation occurs in the Poisson autoregression, (Franke et al., 2012).

Finite sample properties of the proposed tests are explored in a Monte Carlo simulation study in Hudecová (2013b). An application to the US recession data from Figure 1.1 is provided therein as well.

2.3 Dynamic probit model

The model (2.1) can be generalized in various ways. One possibility is to allow for exogenous covariates.

Let $\{Y_t\}$ be a binary time series of interest and let $\{\boldsymbol{x}_t\}$ be a $k \times 1$ vector process of explanatory variables such that $\boldsymbol{x}_t$ includes the constant term for all t. Define the σ-field $\mathcal{F}_t = \sigma\{Y_s, \boldsymbol{x}_{s+1}, s \le t\}$ and assume that the conditional distribution of Y_t given $\mathcal{F}_{t-1}$ is binary with the success probability π_t which depends on $\boldsymbol{x}_t$ and p previous values of the series $Y_{t-1}, \dots, Y_{t-p}$ via the model

$$g(\pi_t) = \sum_{j=1}^{p} \gamma_j Y_{t-j} + \boldsymbol{\beta}^\top \boldsymbol{x}_t, \tag{2.12}$$

where g is a suitable link function. If $g = \Phi^{-1}$ is the quantile function of the standard normal distribution, then the model is referred to as *the binary dynamic response model*, (Kauppi and Saikkonen, 2008; de Jong and Woutersen, 2011). The estimation of the model parameters $\boldsymbol{\theta} = (\gamma_1, \dots, \gamma_p, \boldsymbol{\beta}^\top)^\top$ can be conducted via the maximum conditional likelihood (ML) method, see Kedem and Fokianos (2002) or de Jong and Woutersen (2011), similarly as in the simple model (2.1).

For testing a single change, we introduce the model

$$g(\pi_t) = \begin{cases} \sum_{j=1}^{p} \gamma_j Y_{t-j} + \boldsymbol{\beta}^\top \boldsymbol{x}_t, & t = 1, \dots, m \\ \sum_{j=1}^{p} \gamma_j^* Y_{t-j} + \boldsymbol{\beta}^{*\top} \boldsymbol{x}_t, & t = m+1, \dots, n, \end{cases} \tag{2.13}$$

where $\boldsymbol{\theta} = (\gamma_1, \dots, \gamma_p, \boldsymbol{\beta}^\top) \neq \boldsymbol{\theta}^* = (\gamma_1^*, \dots, \gamma_p^*, \boldsymbol{\beta}^{*\top})$. The test statistics from (2.8)–(2.10) can be easily adapted to the current setup. For instance, we can define

$$W_n = \max_{1 \leq k \leq n} \sqrt{\frac{n}{k(n-k)}} \frac{|\sum_{t=1}^{k} Y_t - \widehat{\pi_t}|}{\sqrt{\frac{1}{n}\sum_{t=1}^{n} \widehat{\pi_t}(1-\widehat{\pi_t})}},$$

where $\widehat{\pi_t}$ is computed from the ML estimator $\widehat{\boldsymbol{\theta}}$ under the null hypothesis of no change. Hudecová (2013a) considered the probit model with $g = \Phi^{-1}$ and proved (along the similar lines as in Hudecová (2013b)) that under some standard regularity conditions the asymptotic distribution of W_n is the same as in Theorem 2.1.

Chapter 3

Count time series

Time series of counts enjoy numerous applications in such diverse fields as business, economics, epidemiology, and traffic analysis. Typically, such data consist of numbers of occurrences of an event over a given time interval, or of counts of objects or individuals.

As already mentioned in Chapter 1, there exist different models which could be considered for a count time series (Davis et al., 2015; McKenzie, 2003). There is an obvious need for a tool which allows one to assess distributional assumptions imposed by the assumed model, i.e., for a goodness-of-fit (GOF) test. The standard approach in constructing GOF tests is to estimate the underlying distribution function or characteristic function and then to construct versions of the Kolmogorov–Smirnov or Cramér–von Mises statistics. However, for count data, the usage of the distribution function is not very convenient due to its non-continuous (piecewise constant) nature, so the GOF tests are typically constructed from different distributional characteristics. Various GOF tests have been suggested in the literature for count time series, (Neumann, 2011; Fokianos and Neumann, 2013; Aleksandrov and Weiss, 2020; Schweer and Weiss, 2014; Weiss et al., 2019; Weiss and Schweer, 2015; Weiss, 2018b).

The time stability of a count time series model has been questioned in the literature as well, which led to various change-point procedures, (e.g., Fokianos and Fried, 2010; Szabó, 2011; Franke et al., 2012; Weiss and Testik, 2009a; Weiss, 2011; Weiss and Testik, 2012, 2015, 2011; Kang and Lee, 2009, 2014; Kirch and Tadjuidje Kamgaing, 2015), or see Lee and Kim (2021) for a recent overview. These articles mostly address the standard problem of parameter change. Relatively popular methods for change–point detection are based on the CUSUM principle (Weiss and Testik, 2009a; Franke et al., 2012; Lee and Kim, 2021), but Wald–type detectors for non–linear INARCH models have also been developed (Doukhan and Kengne, 2015), while likelihood–score criteria applicable to general time series including INARCH models are suggested in Kirch and Tadjuidje Kamgaing (2015).

In the following, a *count distribution* refers to a discrete distribution on $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and a *count variable* is a random variable with such distribution. A count time series is a time series $\{Y_t\}$ such that each Y_t has a count distribution. Furthermore, by a *stationary* time series we mean strictly stationary series.

The structural tests presented in this chapter are based on the *probability generating* *function*, because it is a useful tool for count distributions. Its empirical counterpart has proved to be convenient in various statistical problems, see Nakamura and Pérez-Abreu (1993) for an overview. This concept is introduced in Section 3.1. The considered time series models are briefly revisited in Section 3.2. Section 3.3 deals with the goodness-of-fit tests, while the sequential detection of a change point is studied in Section 3.4.

This chapter is based on results from Hudecová et al. (2015) and Hudecová et al. (2017).

3.1 Probability generating function

If Y is a count random variable then its probability generating function (PGF) is defined as

$$g_Y(u) = \mathsf{E}[u^Y] = \sum_{k=0}^{\infty} u^k \mathsf{P}(Y = k), \quad u \in [0,1].$$

Note that we define $g_Y(u)$ for $u \in [0,1]$, because the power series on the right hand side always converges for $|u| \leq 1$. However, for many distributions it is possible to define g_Y on an interval $(-R, R)$ for some $R \in [1, \infty]$.

The function g_Y uniquely determines the distribution of Y, see (Johnson et al., 2005, Section 1.2.11). Hence, a model for Y can be always specified via the PGF. In the following we use the fact that the PGF of a binomial distribution $\mathsf{Bi}(n,p)$ is $(1-p+pu)^n$ and the PGF of a Poisson distribution with mean λ is $e^{\lambda(u-1)}$. Furthermore, if X and Y are independent, then the PGF of $Z = X + Y$ is

$$g_Z(u) = g_X(u) \cdot g_Y(u),$$

see (Johnson et al., 2005, Section 1.2.11) for further properties.

If $\{Y_t\}$ is an ergodic sequence of stationary variables with marginal PGF $g_Y(u) = \mathsf{E}u^{Y_t}$, then the empirical PGF defined as

$$\widehat{g}_T(u) = \frac{1}{T} \sum_{t=1}^{T} u^{Y_t} \tag{3.1}$$

is a consistent estimator of $g_Y(u)$ as $T \to \infty$, (Rémillard and Theodorescu, 2000). A comparison of the empirical PGF $\widehat{g}_T(u)$ and its theoretical counterpart specified under the assumed model has been successively used in various goodness-of-fit problems, see Baringhaus and Henze (1992) and Nakamura and Pérez-Abreu (1993) for reviews on earlier applications, and see Rémillard and Theodorescu (2000), Meintanis and Bassiakos (2005), Meintanis and Nikitin (2008), Novoa-Muñoz and Jiménez-Gamero (2014), and Novoa-Muñoz and Jiménez-Gamero (2016), among others, for more recent use.

3.2 Count time series models revisited

Recall that $\mathcal{F}_t = \sigma\{Y_s,\ s \leq t\}$ is the information σ-field available at time t. We will revisit INGARCH and INARMA models in Sections 3.2.1 and 3.2.2, respectively.

3.2.1 INARCH and INGARCH models

A nonlinear INGARCH(p,q) model is defined as

$$Y_t|\mathcal{F}_{t-1} \sim F(\lambda_t), \quad \lambda_t = r(Y_{t-1},\dots,Y_{t-p},\lambda_{t-1},\dots,\lambda_{t-q},\boldsymbol{\eta}), \tag{3.2}$$

where $F(\lambda)$ is some count distribution with mean λ, and the function $r : [0,\infty)^{p+q} \times \Theta$ belongs to some specific parametric family of functions $\mathcal{R} = \{r(\cdot,\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k\}$ for some $k \geq 1$. The model where F is Poisson and $\mathcal{R}$ is a family of linear functions was considered by Ferland et al. (2006). The model with non-linear r was considered by Fokianos and Tjøstheim (2009), Fokianos and Tjøstheim (2012), Fokianos (2012). Although the Poisson assumption is by far the most popular specification, different choices for F have been considered as well, see Zhu (2011) and Christou and Fokianos (2014) for a model with conditional negative binomial distribution.

It has been shown that if $p = q = 1$ and r is linear such that $r(\boldsymbol{x},\boldsymbol{\theta}) = \theta_1 + \theta_2 x_1 + \theta_3 x_2$, and if $\theta_i \geq 0$, $i = 1,2,3$, and $\theta_2 + \theta_3 < 1$ and F belongs to the single-parameter exponential family of distributions (that includes the Poisson distribution as a special case), then there exists a strictly stationary and ergodic solution of (3.2), see Davis and Liu (2016). For the Poisson distribution, Ferland et al. (2006) prove that all the moments of the process are finite under the same condition. For an overview of the conditions for strict stationarity and ergodicity regarding other choices of F see, e.g., (Ahmad and Francq, 2016, Section 3).

If $q = 0$ then the model has a purely autoregressive structure and is abbreviated as INARCH(p). We further restrict ourselves to models of order $p = 1$ with the conditional distribution F being Poisson, that is

$$Y_t|\mathcal{F}_{t-1} \sim \mathsf{Po}(\lambda_t), \quad \lambda_t = r(Y_{t-1},\boldsymbol{\theta}). \tag{3.3}$$

If $r(y,\boldsymbol{\theta})$ is assumed to be Lipschitz-continuous in y for all $\boldsymbol{\theta}$ with the Lipschitz constant always strictly smaller than 1, it may be shown that there exists a stationary ergodic solution which is β-mixing with exponential rate, (Neumann, 2011). The simplest case of (3.3) is the linear Poisson autoregression of the first order, where r is linear and

$$Y_t|\mathcal{F}_{t-1} \sim \mathsf{Po}(\lambda_t), \quad \lambda_t = \theta_1 + Y_{t-1}\theta_2 \tag{3.4}$$

with $\boldsymbol{\theta} = (\theta_1,\theta_2)^\top$. It follows from the previously mentioned properties of the more general model that if $\theta_1 > 0$, $\theta_2 \in [0,1)$, then the sequence $\{Y_t\}$ is strictly stationary, ergodic, with all moments finite.

The model parameters of (3.2) can be estimated by the conditional maximum likelihood (ML) method, see Fokianos (2015) and further references therein, or using the Poisson quasi-maximum likelihood, (Ahmad and Francq, 2016).

3.2.2 INAR and INARMA models

A different class of models consists of integer autoregressive moving average (INARMA) models, which mimic the structure of the classical ARMA models, but replace the multiplication sign by the Steutel and van Harn's thinning operator $\circ$ defined in (1.5). Let $\{\varepsilon_t\}$

be a sequence of iid count random variables with distribution G_ε with a finite variance, and let $\alpha_i, \beta_j \in (0,1)$ for $i = 1,\dots,p$ and $j = 1,\dots,q$. The INARMA(p,q) model is defined as

$$Y_t = \sum_{i=1}^{p} \alpha_i \circ Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \beta_j \circ \varepsilon_{t-j}, \tag{3.5}$$

where the Bernoulli variables involved in all the thinning operations are mutually independent and independent of $\{\varepsilon_t\}$. Model (3.5) with $p > 1$ and $q = 0$ was introduced and studied by Du and Li (1991) and since then, many authors have considered various extensions and modifications of model (3.5), see Scotto et al. (2015) for a comprehensive review.

If $p = 1$ and $q = 0$ then the model (3.5) corresponds to the INAR(1) model studied in McKenzie (1985); Al-Osh and Alzaid (1987); Alzaid and Al-Osh (1988), taking the following form

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t = \sum_{i=1}^{Y_{t-1}} U_{t,i} + \varepsilon_t. \tag{3.6}$$

For $\alpha \in (0,1)$ there exists a strictly stationary solution to (3.6) and the law of the innovations $\{\varepsilon_t\}$ uniquely determines the marginal distribution of Y_t as well as the conditional distribution $Y_t|\mathcal{F}_{t-1}$, which is equal to the convolution of binomial $\mathrm{Bi}(Y_{t-1},\alpha)$ and G_ε. In particular, if $\{\varepsilon_t\}$ are iid Poisson then each Y_t has Poisson distribution as well, and this special case has been considered in many applications. In fact, the property that the marginal distribution of the observations is from the same family as the distribution of innovations ε_t characterizes the Poisson law in the context of INAR models, (Al-Osh and Alzaid, 1987). However, some other distributional assumptions have been studied and recommended, (Weiss and Testik, 2009b; Pavlopoulos and Karlis, 2008; Barczy et al., 2010).

The parameters of an INAR(p) model can be estimated by the conditional least squares method or using the conditional maximum likelihood, see (Weiss, 2018a, Chapter 2.2).

3.2.3 Important identities for INAR and INARCH models

Our test statistic for goodness-of-fit testing and change point detection will make use of the properties of the marginal PGF of Y_t following from the model structure.

Assume that $\{Y_t\}$ follows an INAR(1) model (3.6) with parameter $\alpha \in (0,1)$. Since the conditional distribution of $\alpha \circ Y$ given Y is binomial, we can compute the marginal PGF of Y_t as

$$\begin{aligned} g_{Y_t}(u) =& \mathrm{E}\left[u^{Y_t}\right] = \mathrm{E}\left[\mathrm{E}(u^{Y_t}|Y_{t-1})\right] = \mathrm{E}\left[\mathrm{E}(u^{\alpha\circ Y_{t-1}+\varepsilon_t}|Y_{t-1})\right] \\ =& \mathrm{E}\left[(1+\alpha(u-1))^{Y_{t-1}} g_\varepsilon(u)\right] = g_\varepsilon(u) g_{Y_{t-1}}\Big(1+\alpha(u-1)\Big), \end{aligned}$$

where we used the independence of ε_t and Y_{t-1} and denoted as g_ε the PGF of ε_t. If $\{Y_t\}$ is strictly stationary, then we can drop the time index t and write $g_Y = g_{Y_t} = g_{Y_{t-1}}$, which gives the relation

$$g_Y(u) = g_\varepsilon(u) g_Y\Big(1+\alpha(u-1)\Big) \quad \text{for all} \quad u \in [0,1]. \tag{3.7}$$

A similar equality can be derived also for $\{Y_t\}$ following a Poisson INARCH(1) model in (3.3). Under the stationarity assumption, we get

$$g_Y(u) = \mathsf{E}\left[u^{Y_t}\right] = \mathsf{E}\left[\mathsf{E}(u^{Y_t}|Y_{t-1})\right] = \mathsf{E}\left[e^{\lambda_t(u-1)}\right] = \mathsf{E}\left[e^{r(Y_{t-1},\boldsymbol{\theta})(u-1)}\right]. \tag{3.8}$$

For the linear Poisson INARCH(1) in (3.4), we can further write

$$\begin{aligned} g_Y(u) &= \mathsf{E}\left[e^{(\theta_1+\theta_2 Y_{t-1})(u-1)}\right] = e^{\theta_1(u-1)}\mathsf{E}\left[e^{\theta_2(u-1)Y_{t-1}}\right] \\ &= e^{\theta_1(u-1)} g_Y\left(e^{\theta_2(u-1)}\right). \end{aligned} \tag{3.9}$$

3.3 Goodness-of-fit tests based on the probability generating function

We propose to base the GOF test for INARCH and INAR models on a test statistic that is derived as a weighted L_2 distance of two estimators of the marginal PGF, one being fully nonparametric, while the second one is derived under a specified model structure and makes use of identities (3.7) and (3.8) from Section 3.1.

3.3.1 Test statistic

Let $Y_1,\ldots,Y_T$ be given data which come from a stationary series $\{Y_t\}$. Below we formulate the null hypothesis for the models (3.4) and (3.6) from Section 3.2 and propose a suitable test statistic.

Let us start with the INAR(1) model specified by (3.6). Let $\mathcal{G}_\Theta = \{g_\varepsilon(\cdot;\eta);\eta\in\Theta\}$ be some specified family of PGFs of a discrete distribution depending on a finite dimensional parameter $\eta\in\Theta$, where $\Theta\subset\mathbb{R}$ is an open subset. One wishes to test the null hypothesis

$$\mathcal{H}_0^{INAR} : \{Y_t\} \text{ follows model (3.6) for some } \alpha\in(0,1) \text{ and some PGF } g_\varepsilon\in\mathcal{G}_\Theta,$$

against a general alternative that $\mathcal{H}_0^{INAR}$ does not hold. Note that for simplicity we restrict ourselves to the situation where the parameter η is univariate, but the procedure can be very easily adapted to a vectorial parameter $\boldsymbol{\eta}$, if the assumptions about smoothness of g_ε with respect to η stated in assumption (A.3) in the next section are modified accordingly.

Similarly, we formulate a null hypothesis which states that the data come from a linear Poisson INARCH(1) model

$$\mathcal{H}_0^{PAR}: \ \{Y_t\} \text{ follows model (3.4) for some } \boldsymbol{\theta}\in A,$$

where $A = \{(\theta_1,\theta_2)^\top : \theta_1 > 0,\ \theta_2\in(0,1)\}$, against a general alternative.

We propose a goodness-of-fit test statistic which is based on a comparison of two estimators of the marginal PGF of Y_t. The first one is the empirical PGF defined in (3.1) which is a fully non–parametric estimate of g_Y, valid for a broad class of stationary processes. The second estimator $\widetilde{g}_T(u)$ is semiparametric derived from relation (3.7) for INAR and from (3.9) for the linear Poisson INARCH, so it is valid only under the null hypothesis. The two estimators are then compared in the test statistic

$$S_T = T\int_0^1 \Big(\widehat{g}_T(u) - \widetilde{g}_T(u)\Big)^2 w(u)du, \tag{3.10}$$

where $w : (0,1) \to [0,\infty)$ is a nonnegative weight function.

It remains to provide the formula for the semiparametric estimator $\widetilde{g}_T(u)$. Under $\mathcal{H}_0^{INAR}$, let $g_\varepsilon(\cdot;\eta)$ be the PGF of ε_t and let $\widehat{\alpha}$ and $\widehat{\eta}$ be suitable estimators of α and η respectively constructed from $Y_1,\ldots,Y_T$ under $\mathcal{H}_0^{INAR}$ (where we for simplicity suppress the index T). Then it follows from (3.7) that the estimator $\widetilde{g}_T$ can be defined as

$$\widetilde{g}_T(u) = g_\varepsilon(u;\widehat{\eta})\widehat{g}_T\Big(1+\widehat{\alpha}(u-1)\Big), \quad u \in [0,1]. \tag{3.11}$$

Analogously, for model (3.4), a natural semiparametric estimate of the marginal PGF follows from (3.9). Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be suitable estimators of θ_1 and θ_2, respectively. Define

$$\widetilde{g}_T(u) = e^{\widehat{\theta}_1(u-1)}\widehat{g}_T(e^{\widehat{\theta}_2(u-1)}).$$

Even though the test statistic in (3.10) is expressed in terms of a definite integral, its computation can be simplified, see Hudecová et al. (2015) for more details. The next sections provide the limit behavior of the test statistic S_T for the sample size $T \to \infty$ for INAR(1) and INARCH(1).

Remark 3.1. The proposed procedure can be straightforwardly generalized to non-linear or non-Poisson INARCH models and to models of higher orders, see Hudecová et al. (2015). A generalization to multivariate count time series models is considered in Hudecová et al. (2021). A goodness-of-fit test for an INGARCH model with exogenous covariates is treated in Hudecová et al. (2024), where the PGF is combined with the characterization from Bierens (1982).

3.3.2 Asymptotic results and a bootstrap test for INAR

If not stated explicitly, all convergences and O_P and o_p notions are meant for $T \to \infty$.

Recall that under $\mathcal{H}_0^{INAR}$ the PGF of ε_t is $g_\varepsilon(u) \in \mathcal{G}_\Theta$, i.e., $g_\varepsilon(\cdot;\eta)$ is specified up to a parameter $\eta \in \Theta$, with Θ being an open set in $\mathbb{R}$. Consider the following assumptions:

(A.1) Let $\{Y_t\}$ be the stationary solution to (3.6) with the true parameter α_0 and let $\{\varepsilon_t\}$ be a sequence of iid count random variables with a finite variance and PGF g_ε.

(A.2) Let $g_\varepsilon \in \mathcal{G}_\Theta$ and $g_\varepsilon = g_\varepsilon(\cdot;\eta_0)$, $\eta_0 \in \Theta$, where Θ is an open subset of $\mathbb{R}$.

(A.3) The first partial derivative of $g_\varepsilon(u;\eta)$ with respect to η exists for all $u \in [0,1]$ and there exist constants $0 < D_j < \infty$, $j = 1,2,3$, and a measurable function $v : [0,1] \to (0,\infty)$ such that

$$\left|\frac{\partial g_\varepsilon}{\partial \eta}(u;\eta) - \frac{\partial g_\varepsilon}{\partial \eta}(u;\eta_0)\right| \le D_1|\eta-\eta_0|v(u) \quad \text{for all } u \in [0,1],\ |\eta-\eta_0| \le D_2,$$

and

$$\left|\frac{\partial g_\varepsilon(u;\eta)}{\partial \eta}\right| \le D_3 v(u), \quad \text{for all } u \in [0,1],\ |\eta-\eta_0| \le D_2.$$

(A.4) The weight function $w:(0,1)\to[0,\infty)$ satisfies $0<\int_0^1 w(u)du<\infty$ and $\int_0^1 w(u)v^2(u)du<\infty$, where v is from (A.3).

(A.5) Let $\widehat{\vartheta}=(\widehat{\alpha},\widehat{\eta})^\top$ be an estimator of the true value $\vartheta_0=(\alpha_0,\eta_0)^\top$ satisfying

$$\sqrt{T}(\widehat{\vartheta}-\vartheta_0)=\frac{1}{\sqrt{T}}\sum_{t=1+q}^{T}\ell(\mathbf{Y}_{t-q};\vartheta_0)+o_P(1),$$

for some fixed $q\geq 1$, where $\mathbf{Y}_{t-q}=(Y_t,\ldots,Y_{t-q})^\top$, and $\ell=(\ell_1,\ell_2)^\top$ is such that $\ell_j(\mathbf{Y}_{t-q};\vartheta_0)$, $j=1,2$, are martingale difference sequences with finite variances.

The assumption (A.1) states that $\mathcal{H}_0^{INAR}$ holds. The assumption (A.3) requires some smoothness of the PGF g_ε. It is easy to verify that all the requirements are fulfilled for the baseline Poisson distribution, where $g_\varepsilon(u;\eta)=e^{\eta(u-1)}$. The assumption (A.5) requires a "reasonable" estimator of the model parameters. Note that the conditional least squares estimator satisfies (A.5).

Theorem 3.1. *Let assumptions (A.1)–(A.5) be satisfied. Then*

$$S_T\xrightarrow{D}\int_0^1\mathcal{Z}^2(u)w(u)du,$$

where $\left\{\mathcal{Z}(u);\ u\in[0,1]\right\}$ *is a zero-mean Gaussian process with the covariance structure*

$$\begin{aligned}\mathsf{E}\mathcal{Z}(u_1)\mathcal{Z}(u_2)=\mathsf{E}\Big[&u_1^{Y_{q+1}}-\Big(1+\alpha_0(u_1-1)\Big)^{Y_q}g_\varepsilon(u_1;\eta_0)+h_1(u_1;\alpha_0,\eta_0)\ell_1(\mathbf{Y}_1;\alpha_0,\eta_0)\\&+h_2(u_1;\alpha_0,\eta_0)\ell_2(\mathbf{Y}_1;\alpha_0,\eta_0)\Big]\\\times\Big[&u_2^{Y_{q+1}}-\Big(1+\alpha_0(u_2-1)\Big)^{Y_q}g_\varepsilon(u_2;\eta_0)+h_1(u_2;\alpha_0,\eta_0)\ell_1(\mathbf{Y}_1;\alpha_0,\eta_0)\\&+h_2(u_2;\alpha_0,\eta_0)\ell_2(\mathbf{Y}_1;\alpha_0,\eta_0)\Big],\end{aligned}$$

where $u_1,u_2\in[0,1]$, *and*

$$h_1(u;\alpha,\eta)=\frac{\partial g_Y(1+\alpha(u-1))}{\partial\alpha}g_\varepsilon(u;\eta),\quad h_2(u;\alpha,\eta)=g_Y(1+\alpha(u-1))\frac{\partial g_\varepsilon(u;\eta)}{\partial\eta}.$$

Theorem 3.1 expresses the limiting distribution of S_T as a distribution of an integral of a Gaussian process with some specified covariance structure. This means that the limiting distribution is equal to an infinite weighted sum of independent χ_1^2 variables, where the weights depend on the unknown parameters in a non-trivial way. Hence, there is no explicit form for the limiting distribution function and, in addition, this distribution function depends on unknown quantities including the true value of the parameter ϑ. Consequently, the result from Theorem 3.1 is not directly applicable for the purpose of approximating critical values and actually performing the test. Nevertheless, when a consistent estimator of the covariance structure is available, we can use it to obtain an approximation of the limiting distribution.

Alternatively, some resampling methods can be applied in order to carry out the test procedure and to compute the critical values. In the current setup it is quite natural to use the parametric bootstrap as a resampling scheme because it reflects all aspects of the underlying model, and has been put on a firm theoretical basis both with iid data, (Genest and Rémillard, 2008), as well as with data involving dependence, (Leucht and Neumann, 2013). The procedure follows straightforwardly as described in Algorithm 1.

Algorithm 1 Parametric bootstrap for GOF for $\mathcal{H}_0^{INAR}$.

Input: Data $Y_1, \dots, Y_T$.

compute the estimator $\widehat{\vartheta} = (\widehat{\alpha}, \widehat{\eta})^\top$ and the test statistic S_T
for $b = 1$ to B **do**
  generate $\varepsilon_1^{*b}, \dots, \varepsilon_T^{*b}$ independently from a distribution with PGF $g_\varepsilon(u; \widehat{\eta})$
  generate sequentially pseudo–observations $Y_1^{*b}, \dots, Y_T^{*b}$, as $Y_1^{*b} = \varepsilon_1^{*b}$ and
  for $t = 2$ to T **do**
    generate iid variables $U_{t,i}^*$, $i = 1, \dots, Y_{t-1}^{*b}$ from a binary distribution $\mathrm{B}(\widehat{\alpha})$
    compute

$$Y_t^{*b} = \sum_{i=1}^{Y_{t-1}^{*b}} U_{t,i}^* + \varepsilon_t^{*b}$$

  end for
  fit the model (3.6) for data Y_t^{*b}, $t = 1, \dots, T$, to obtain the estimate $\widehat{\vartheta}^{*b}$
  compute the test statistic S_T^{*b}
end for
compute the p-value as

$$p = \frac{1}{B+1} \sum_{b=1}^{B} \mathrm{I}[S_T^{*b} \geq S_T]$$

Output: p-value of significance of $\mathcal{H}_0^{INAR}$

Consider now the behavior of the test statistic under alternatives, where the PGF of ε_t is not correctly specified, that is $g_\varepsilon \notin \mathcal{G}_\Theta$. Assume that

(A.6) Let $(\widehat{\alpha}, \widehat{\eta})^\top$ be an estimator such that

$$(\widehat{\alpha}, \widehat{\eta})^\top \xrightarrow{P} (\alpha_0, \eta_A)^\top \tag{3.12}$$

for $\alpha_0 \in (0,1)$ and for some $\eta_A \in \Theta$.

Theorem 3.2. *Let (A.1), (A.4) and (A.6) be satisfied and let $g_\varepsilon(u;\eta)$ be continuous in η for all $u \in [0,1]$. Then*

$$\frac{S_T}{T} \xrightarrow{P} \int_0^1 \Big[g_Y(1 + \alpha_0(u-1))(g_\varepsilon(u) - g_\varepsilon(u;\eta_A)) \Big]^2 w(u) \mathrm{d}u. \tag{3.13}$$

The right–hand side of (3.13) is positive unless the true innovation PGF g_ε coincides with the PGF $g_\varepsilon(\cdot;\eta_A)$ postulated under the null hypothesis $\mathcal{H}_0^{INAR}$. This and the uniqueness of the PGF implies the consistency of the test which rejects the null hypothesis for large values of the test statistic S_T under fixed alternatives. However, we should point out that despite the fact that the formulation of the alternative adopted here focuses exclusively on the innovation PGF, the construction of the test statistic is based not only on the specification of the PGF but on the entire INAR model. Indeed, Monte Carlo simulations conducted in Hudecová et al. (2015) reveal that the test detects other departures from the null hypothesis as well.

3.3.3 Asymptotic results and a bootstrap test for INARCH model

In order to test the null hypothesis $\mathcal{H}_0^{PAR}$ of a linear Poisson INARCH(1) model in (3.4), consider the following assumptions:

(B.1) Let $\{Y_t\}$ be the stationary solution to (3.4) with the true parameter $\boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20})^\top \in A$, where $A = \{(\theta_1, \theta_2)^\top : \theta_1 > 0,\ \theta_2 \in (0,1)\}$.

(B.2) Let $(\widehat{\theta}_1, \widehat{\theta}_2)^\top$ be an estimator of $(\theta_1, \theta_2)^\top$ satisfying

$$\sqrt{T}(\widehat{\theta}_j - \theta_{j0}) = \frac{1}{\sqrt{T}} \sum_{t=q+1}^{T} \ell_j(\mathbf{Y}_{t-q}; \theta_{10}, \theta_{20}) + o_P(1), \quad j = 1, 2, \tag{3.14}$$

where $\ell_j(\mathbf{Y}_{t-q}; \theta_{10}, \theta_{20})$, $j = 1, 2$, for fixed $q \geq 1$, are martingale difference sequences with finite variances.

(B.3) Assume that the weight function $w \geq 0$ satisfies $\int_0^1 w(u)du < \infty$.

Theorem 3.3. *Let (B.1)–(B.3) hold. Then*

$$S_T \xrightarrow{D} \int_0^1 \mathcal{V}^2(u) w(u) du,$$

where $\left\{\mathcal{V}(u);\ u \in [0,1]\right\}$ *is a zero-mean Gaussian process with the covariance structure*

$$\begin{aligned}
\mathsf{E}\mathcal{V}(u_1)\mathcal{V}(u_2) = \mathsf{E}\Big[& u_1^{Y_{q+1}} - \exp\Big\{(\theta_{10} + \theta_{20} Y_q)(u_1 - 1)\Big\} + r_1(u_1; \theta_{10}, \theta_{20})\ell_1(\mathbf{Y}_1; \theta_{10}, \theta_{20}) \\
&+ r_2(u_1; \theta_{10}, \theta_{20})\ell_2(\mathbf{Y}_1; \theta_{10}, \theta_{20})\Big] \\
\times \Big[& u_2^{Y_{q+1}} - \exp\Big\{(\theta_{10} + \theta_{20} Y_q)(u_2 - 1)\Big\} + r_1(u_2; \theta_{10}, \theta_{20})\ell_1(\mathbf{Y}_1; \theta_{10}, \theta_{20}) \\
&+ r_2(u_2; \theta_{10}, \theta_{20})\ell_2(\mathbf{Y}_1; \theta_{10}, \theta_{20})\Big],
\end{aligned}$$

where $u_1, u_2 \in [0,1]$, *and*

$$\begin{aligned}
r_1(u; \theta_1, \theta_2) =& \mathbb{E}\Big[\exp\Big\{(\theta_1 + \theta_2 Y_q)(u - 1)\Big\}(u - 1)\Big], \\
r_2(u; \theta_1, \theta_2) =& \mathbb{E}\Big[Y_q \exp\Big\{(\theta_1 + \theta_2 Y_q)(u - 1)\Big\}(u - 1)\Big].
\end{aligned}$$

Similarly as for the INAR case, we recommend to evaluate the significance of S_T using a parametric bootstrap test, which proceeds here naturally, as described in Algorithm 2.

Algorithm 2 Parametric bootstrap for GOF for $\mathcal{H}_0^{PAR}$.

Input: Data $Y_1, \ldots, Y_T$.

1: compute the estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2)^\top$ and the test statistic S_T

2: **for** $b = 1$ to B **do**

3: generate sequentially pseudo–observations $Y_1^{*b}, \ldots, Y_T^{*b}$ as $Y_1^{*b} = Y_1$ and

4: **for** $t = 2$ to T **do**

5: generate Y_t^{*b} from a Poisson distribution with mean

$$\widehat{\theta}_1 + \widehat{\theta}_2 Y_{t-1}^{*b}$$

6: **end for**

7: fit the model (3.4) for data Y_t^{*b}, $t = 1, \ldots, T$, to obtain the estimate $\widehat{\boldsymbol{\theta}}^{*b}$

8: compute the test statistic S_T^{*b}

9: **end for**

10: compute the p-value as

$$p = \frac{1}{B+1} \sum_{b=1}^{B} \mathrm{I}[S_T^{*b} \geq S_T]$$

Output: p-value of significance of $\mathcal{H}_0^{PAR}$

3.4 Testing for a change

The objective of this section is to construct a test statistic for a *sequential detection*, introduced in Section 1.4, of structural breaks in INAR and Poisson INARCH time series. We again employ PGF as our main tool in the change–detector statistics. Our motivation stems from earlier results obtained for classical continuous–type data based on the characteristic function. In particular, Hušková and Meintanis (2006) and Hlávka et al. (2016) compare classical approaches for change–point detection, such as the empirical likelihood, Kolmogorov–Smirnov, Cramér–von Mises and CUSUM methods, to corresponding methods based on the empirical characteristic function, and report favorable results for the latter methods for iid as well as time series data.

Within the count time series models from Section 3.2, we can either consider a *parametric change* or a more general *nonparametric change* in the model. In the parametric approach, we often assume a specific parametric form of the underlying distribution and test for a single abrupt change in the model parameters. More generally, within the nonparametric framework, we can test for a change in the whole, possibly unspecified, distribution.

In the following, G_Y stands for the cumulative distribution function (cdf) of a count random variable Y and g_Y is its probability generating function (PGF).

We illustrate the situation on the INAR(1) model: Assume that $\{Y_t\}$ is generated by the INAR(1) model (3.6) with a sequence $\{\varepsilon_t\}$ with a finite variance such that ε_t has cdf $G_{\varepsilon,t}$, and that the variables $U_{t,i}$ follow the Bernoulli distribution $\mathsf{B}(\alpha_t)$.

(i) We can consider the null hypothesis

$$\mathcal{H}_0 : \alpha_t = \alpha_0,\ G_{\varepsilon,t} = G_\varepsilon^{(0)},\ \forall t, \tag{3.15}$$

against the alternative

$$\mathcal{H}_1 : \alpha_t = \alpha_0,\ G_{\varepsilon,t} = G_\varepsilon^{(0)},\ t \le T + t_0; \quad \alpha_t = \alpha^0,\ G_{\varepsilon,t} = G_\varepsilon^{(1)},\ t > T + t_0,$$

for some unknown $(\alpha_0, G_\varepsilon^{(0)}) \neq (\alpha^0, G_\varepsilon^{(1)})$, and an unknown change point t_0.

(ii) Within the parametric framework, we assume that the PGF of ε_t belongs to a given parametric family $\mathcal{G}_\Theta = \{g(\cdot;\eta) : \eta \in \Theta\}$, where $\Theta \subset \mathbb{R}$ is an open set so that $g_{\varepsilon_t}(u) = g(u, \eta_t)$. The null hypothesis of no change

$$\mathcal{H}_0 : (\alpha_t, \eta_t) = (\alpha_0, \eta_0), \quad 1 \le t \le T,$$

is tested against the alternative

$$\mathcal{H}_1 : (\alpha_t, \eta_t) = (\alpha_0, \eta_0), \quad 1 \le t \le T + t_0, \quad (\alpha_t, \eta_t) = (\alpha^0, \eta^0), \quad t > T + t_0,$$

for some unknown $(\alpha_0, \eta_0) \neq (\alpha^0, \eta^0)$ and an unknown change point t_0.

It is clear that (ii) is a special case of (i). Here, we focus on the more general type of changes in (i), for both the INAR and INARCH models. Note that parametric changes in the same setup were considered in Hudecová et al. (2015) and Hudecová et al. (2016).

We consider the sequential (online) detection with a monitoring period mT for some $m > 0$, see Section 1.4. We need to construct a test statistic $\widehat{\Delta}_{T,t}$ that would be sensitive to violations of a specified $\mathcal{H}_0$. The null hypothesis will be then rejected if the detector statistic $D_{T,t,\gamma}$ defined in (1.12) exceeds the threshold $c_{\alpha,m}$ computed such that the resulting test keeps the prescribed level α asymptotically, i.e., (1.15) holds.

3.4.1 Change point statistics for INAR

Let $\{Y_t\}$ be generated by the INAR(1) model in (3.6) with $U_{t,i}$ following the Bernoulli distribution $\mathrm{B}(\alpha)$ (no change) and with a sequence $\{\varepsilon_t\}$ of independent count random variables with a finite variance and cdfs $G_{\varepsilon,t}$. Consider

$$\mathcal{H}_0^{INAR} : G_{\varepsilon,t} = G_\varepsilon^{(0)},\ \forall t, \tag{3.16}$$

against the alternative

$$\mathcal{H}_1^{INAR} : G_{\varepsilon,t} = G_\varepsilon^{(0)},\ t \le T + t_0;\ \ G_{\varepsilon,t} = G_\varepsilon^{(1)},\ t > T + t_0,$$

for some unknown innovation distributions $G_\varepsilon^{(0)} \neq G_\varepsilon^{(1)}$ and an unknown change point t_0. Within this setup we concentrate on changes in $G_{\varepsilon,t}$ but it will also be seen that the suggested test is sensitive even to a change due to a time–varying Bernoulli parameter α_t of the variables $U_{t,i}$. Moreover, we will prove consistency of our test statistic under a more general alternative (see the formulation in (C.4) later).

Recall that under $\mathcal{H}_0^{INAR}$ we can assume that $\{Y_t\}$ is stationary, and let $g_{\varepsilon,0}$ and $g_{Y,0}$ be the PGF of ε_t and Y_t, respectively. It follows from (3.7) that

$$g_{\varepsilon,0}(u) = \frac{g_{Y,0}(u)}{g_{Y,0}(1+\alpha(u-1))}, \quad u \in [0,1]. \tag{3.17}$$

Let $\widehat{\alpha} = \widehat{\alpha}_T$ (we will again suppress the subscript T for brevity) be an estimator of the Bernoulli parameter α constructed from the training data $\{Y_t,\ t = 1,\dots,T\}$ without imposing any specific structure on $g_{\varepsilon,0}$ (e.g., $\widehat{\alpha}$ is a moment estimator). Then, based on the training data, the estimator of $g_{\varepsilon,0}(u)$ is given by

$$\widehat{g}_{\varepsilon,T}(u) = \frac{\widehat{g}_{Y,T}(u)}{\widehat{g}_{Y,T}(1+\widehat{\alpha}(u-1))}, \tag{3.18}$$

where $\widehat{g}_{Y,T}(u)$ is the empirical PDF defined in (3.1). An analogous estimator computed from data $Y_{T+1},\dots,Y_{T+t}$ is given by

$$\widehat{g}_{\varepsilon,T,T+t}(u) = \frac{\widehat{g}_{Y,T,T+t}(u)}{\widehat{g}_{Y,T,T+t}(1+\widehat{\alpha}(u-1))} \tag{3.19}$$

where

$$\widehat{g}_{Y,T,T+t}(u) = \frac{1}{t}\sum_{\tau=T+1}^{T+t} u^{Y_\tau}.$$

Under $\mathcal{H}_0^{INAR}$, the difference

$$\widehat{g}_{\varepsilon,T,T+t}(u) - \widehat{g}_{\varepsilon,T}(u)$$

should be small for all $u \in [0,1]$, while under $\mathcal{H}_1^{INAR}$ it is expected to be large at some u. See that

$$\begin{aligned}|\widehat{g}_{\varepsilon,T}(u) - \widehat{g}_{\varepsilon,T,T+t}(u)| &= \frac{|\widehat{g}_{Y,T}(\widehat{u}_\alpha)\widehat{g}_{Y,T,T+t}(u) - \widehat{g}_{Y,T}(u)\widehat{g}_{Y,T,T+t}(\widehat{u}_\alpha)|}{\widehat{g}_{Y,T}(\widehat{u}_\alpha)\widehat{g}_{Y,T,T+t}(\widehat{u}_\alpha)} \\ &\geq |\widehat{g}_{Y,T}(\widehat{u}_\alpha)\widehat{g}_{Y,T,T+t}(u) - \widehat{g}_{Y,T}(u)\widehat{g}_{Y,T,T+t}(\widehat{u}_\alpha)|,\end{aligned}$$

where

$$\widehat{u}_\alpha = 1 + \widehat{\alpha}(u-1). \tag{3.20}$$

Hence, we propose a test statistic defined by the weighted integral

$$\widehat{\Delta}_{T,t} = \int_0^1 \widehat{\delta}^2_{T,t}(u) w(u) \mathrm{d}u, \tag{3.21}$$

where $w : (0,1) \to (0,\infty)$ denotes a nonnegative weight function and

$$\widehat{\delta}_{T,t}(u) = \frac{t}{T^{1/2}}\Big[\widehat{g}_{Y,T}(\widehat{u}_\alpha)\widehat{g}_{Y,T,T+t}(u) - \widehat{g}_{Y,T}(u)\widehat{g}_{Y,T,T+t}(\widehat{u}_\alpha)\Big]. \tag{3.22}$$

If $\widehat{\delta}_{T,t}(u)$ is large, then also $|\widehat{g}_{\varepsilon,T}(u) - \widehat{g}_{\varepsilon,T,T+t}(u)|$ is large, which happens if there is a change in the model. Hence, large values of $\widehat{\Delta}_{T,t}$ indicate a violation of $\mathcal{H}_0^{INAR}$.

3.4.2 Change point statistics for INARCH

Consider a nonlinear INARCH(1) model from (3.2), where we assume that

$$Y_t|\mathcal{F}_{t-1} \sim F_{Y,t} = F_t(\lambda_t), \quad \lambda_t = r_t(Y_{t-1}, \boldsymbol{\theta}), \tag{3.23}$$

where $\boldsymbol{\theta} \in A$ for A a compact subset of $\mathbb{R}^d$, $d \in \mathbb{N}$. The null hypothesis of a Poisson INARCH(1) without a change point is stated as

$$\mathcal{H}_0^{PAR} : F_{Y,t} = F^{(0)}(\lambda_t), \quad \lambda_t = r_0(Y_{t-1}, \boldsymbol{\theta}_0), \quad \forall t, \tag{3.24}$$

where r_0 is known up to finitely many parameters and $F^{(0)}(\lambda)$ denotes the Poisson distribution with parameter λ. The null hypothesis $\mathcal{H}_0^{PAR}$ will be tested against the alternative

$$\mathcal{H}_1^{PAR} : F_{Y,t} = \begin{cases} F^{(0)}(\lambda_t), & \lambda_t = r_0(Y_{t-1}, \boldsymbol{\theta}_0), \quad t \leq T + t_0; \\ F^{(1)}(\lambda_t), & \lambda_t = r_1(\mathcal{F}_{t-1}) \qquad t > T + t_0, \end{cases}$$

where $F^{(1)}$ is an unspecified count distribution such that $F^{(1)} \neq F^{(0)}$, and the change point t^0 is unknown.

The alternative $\mathcal{H}_1^{PAR}$ covers as particular cases, a change in the parameter, say from $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}^0$ (both considered unknown in our setting), change in the regression function, from a fixed known $r_0(\cdot,\cdot)$ to an arbitrary regression function $r^0(\cdot,\cdot)$, as well as a structural change according to which the conditional law $F_{Y,t}$ changes from a Poisson to a non–Poisson specification.

Construction of the detector test statistic for $\mathcal{H}_0^{PAR}$ is based on the relationship (3.8). Recall that under $\mathcal{H}_0^{PAR}$ it holds that

$$g_Y(u) = \mathsf{E}e^{r_0(Y_{t-1}, \boldsymbol{\theta}_0)(u-1)}.$$

Similarly as in Section 3.3.1, we can construct a semi-parametric estimator of g_Y and compare it with the fully nonparametric empirical PGF. Assume that $\{Y_t\}$ is a stationary and ergodic solution of $\mathcal{H}_0^{PAR}$. If $\boldsymbol{\theta}_0$ was known, then the following four estimators would be consistent for $g_Y(u)$ as $T, t \to \infty$:

$$\text{(E1)} \quad \frac{1}{T}\sum_{j=1}^{T} u^{Y_j}, \qquad \text{(E2)} \quad \frac{1}{T}\sum_{j=2}^{T} e^{r_0(Y_{j-1}, \boldsymbol{\theta}_0)(u-1)},$$

$$\text{(E3)} \quad \frac{1}{t}\sum_{j=T+1}^{T+t} u^{Y_j}, \qquad \text{(E4)} \quad \frac{1}{t}\sum_{j=T+1}^{T+t} e^{r_0(Y_{j-1}, \boldsymbol{\theta}_0)(u-1)}.$$

Under the alternative, only the first three estimators (E1)–(E3) are consistent. Therefore, our test statistic is based on the comparison of the difference between the nonparametric and semiparametric estimators in the training data and in the monitoring sample. Let $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_T$ be an estimator of $\boldsymbol{\theta}_0$ constructed from the training data. Define

$$\begin{aligned} \widehat{\delta}_{T,t}(u; \widehat{\boldsymbol{\theta}}_T) = {} & \frac{t}{\sqrt{T}} \frac{1}{t} \sum_{j=T+1}^{T+t} \left(u^{Y_j} - \exp\{r_0(Y_{j-1}, \widehat{\boldsymbol{\theta}})(u-1)\} \right) \\ & - \frac{t}{\sqrt{T}} \frac{1}{T} \sum_{j=2}^{T} \left(u^{Y_j} - \exp\{r_0(Y_{j-1}, \widehat{\boldsymbol{\theta}})(u-1)\} \right) \end{aligned} \tag{3.25}$$

and

$$\widehat{\Delta}_{T,t} = \int_0^1 \widehat{\delta}^2_{T,t}(u;\widehat{\boldsymbol{\theta}})w(u)\mathrm{d}u, \tag{3.26}$$

where $w : (0,\infty] \to (0,\infty)$ is a nonnegative weight function. Large values of $\widehat{\Delta}_{T,t}$ indicate violations of $\mathcal{H}_0^{PAR}$.

3.4.3 Asymptotics and a bootstrap test

In view of (1.15), we have to study the limit behavior of

$$M_{T,\gamma} := \max_{1\le t\le mT} \frac{1}{q_\gamma^2(t/T)} \int_0^1 \widehat{\delta}^2_{T,t}(u)w(u)\mathrm{d}u, \tag{3.27}$$

where $\widehat{\delta}^2_{T,t}(u)$ is defined in (3.22) for the INAR model and in (3.25) for the Poisson INARCH model. The limit behavior is considered always for $T \to \infty$ and m fixed.

Consider first the INAR model. For the behavior under the null hypothesis $\mathcal{H}_0^{INAR}$ from (3.16), we state the following assumptions:

(C.1) Let $\{Y_t\}$ be the stationary solution to (3.6) with $\{\varepsilon_t\}$ iid random variables with a finite variance.

(C.2) Let $\widehat{\alpha}$ be an estimator of parameter α based on training data $Y_1,\dots,Y_T$ satisfying

$$\sqrt{T}(\widehat{\alpha}-\alpha) = O_P(1). \tag{3.28}$$

(C.3) The weight function satisfies $w \ge 0$ and $0 < \int_0^1 w(u)\mathrm{d}u < \infty$.

Theorem 3.4. *Let (C.1)–(C.3) hold and let $M_{T,\gamma}$ be defined as in (3.27) with $\widehat{\delta}_{T,t}(u)$ being defined by (3.22) and $q_\gamma(\cdot)$, $\gamma \in [0,1/2)$ in (1.13). Then*

$$M_{T,\gamma} \xrightarrow{D} \sup_{0<s<m/(m+1)} \frac{1}{s^{2\gamma}} \int_0^1 V^2(s,u)\,[\mathrm{E}(1+\alpha(u-1))^{Y_1}]^2 w(u)\mathrm{d}u, \tag{3.29}$$

where $\Big\{V(s,t);\, s \in (0,m/(m+1)),\, u \in (0,1)\Big\}$ is a zero-mean Gaussian process with the covariance structure

$$\mathrm{E}\Big(V(s_1,u_1), V(s_2,u_2)\Big) = \min(s_1,s_2)\sigma(u_1,u_2)$$

with

$$\sigma(u_1,u_2) = E\Big[\Big(u_1^{Y_2} - E(u_1^{Y_2}|Y_1)\Big)\Big(u_2^{Y_2} - E(u_2^{Y_2}|Y_1)\Big)\Big],$$

where $E(u^{Y_2}|Y_1)$ is the conditional PGF of Y_2 given Y_1.

Unfortunately, there is no explicit form for the asymptotic distribution in (3.29) and its quantiles cannot be computed analytically. However. if $\widehat{\sigma}(u_1,u_2)$ is a consistent estimator of $\sigma(u_1,u_2)$, then it is possible to simulate the corresponding Gaussian process and compute the random variable on the right hand side of (3.29). Hence, an approximation for the desired critical value $c_{\alpha,m}$ can be computed numerically using a Monte Carlo method.

To the best of our knowledge, however, this approach has never been realized in the current context as it is very computationally demanding and the quality of the approximation depends strongly on the properties of $\widehat{\sigma}(u_1, u_2)$. On the other hand, it is far more straightforward to approximate the asymptotic null distribution of $M_{T,\gamma}$ by a proper version of bootstrap.

Since it is not possible to use a parameteric bootstrap under $\mathcal{H}_0^{INAR}$ (because $G_\varepsilon^{(0)}$ is not specified), we propose the following resampling procedure, whose validity is justified in Theorem 3.5. The procedure is motivated by the following expression for $\widehat{\delta}_{T,t}(u)$ in (3.21). See that it follows from (3.22) that

$$\widehat{\delta}_{T,t}(u) = \frac{1}{T^{3/2}}\Big(\sum_{j=1}^{T} \widehat{u}_\alpha^{Y_j} \sum_{s=T+1}^{T+t} u^{Y_s} - \sum_{j=1}^{T} u^{Y_j} \sum_{s=T+1}^{T+t} \widehat{u}_\alpha^{Y_s}\Big),$$

where $\widehat{u}_\alpha$ is defined in (3.20). Denote

$$B_{T,t}(u) = \widehat{u}_\alpha^{Y_t}, \quad t = 1,\dots,T(m+1), \tag{3.30}$$

and

$$A_{T,t}(u) = u^{Y_t} - \widehat{u}_\alpha^{Y_{t-1}} \widehat{g}_{\varepsilon,T}(u), \quad t = 2,\dots,T(m+1). \tag{3.31}$$

Then

$$A_{T,t}(u) = u^{Y_t} - \widehat{g}_{\varepsilon,T}(u) B_{T,t-1}(u).$$

Set

$$\widetilde{\delta}_{T,t}(u) = \frac{1}{T\sqrt{T}}\left(\sum_{s=T+1}^{T+t} A_{T,s}(u) \sum_{j=1}^{T} B_{T,j}(u) - \sum_{j=2}^{T} A_{T,j}(u) \sum_{s=T+1}^{T+t} B_{T,s}(u)\right). \tag{3.32}$$

Lemma 3.1. *Under the assumptions of Theorem 3.4, the limiting distribution of $M_{T,\gamma}$ is the same as the limiting distribution of $\widetilde{M}_{T,\gamma}$ which is defined by (3.27) with $\widehat{\delta}_{T,t}$ replaced with $\widetilde{\delta}_{T,t}(u)$.*

The idea of our resampling method is to generate iid pairs of consecutive observations (Y_{t-1}, Y_t) and to use formulas (3.30)–(3.32) for the computation of the bootstrap versions of the test statistic, as summarized in Algorithm 3. Theorem 3.5 justifies the proposed method.

Theorem 3.5. *Let $\widetilde{M}^{*b}$, $b = 1,\dots,B$, be computed as in Algorithm 3. Under the assumptions of Theorem 3.4, it holds that*

$$P(\widetilde{M}_{T,\gamma}^{*b} \le x | Y_1,\dots,Y_T) - P(M_{T,\gamma} \le x) \xrightarrow{P} 0$$

for all $x \in \mathbb{R}$ as $T \to \infty$.

The assertion implies that the proposed bootstrap based on historical observations provides an asymptotically correct approximation for the distribution of the studied test statistic under the null hypothesis and, therefore, provides the desired approximation for critical values.

We also establish the limit behavior of the proposed procedure under the following alternatives:

Algorithm 3 Bootstrap procedure for a change point detection in INAR model.

Input: Training data $Y_1,\dots,Y_T$, observations $\{Y_k\}_{k=T+1}^{mT}$. Tuning parameter $\gamma \in (0,1/2]$.

compute $\widehat{\alpha}$ and $\widehat{g}_{\varepsilon,T}(u)$ from (3.18)
compute $\widetilde{M}_{T,\gamma}$
for $b=1$ to B **do**
  generate iid bootstrap pairs (Y_t^{*b}, Y_{t-1}^{*b}), $t=2,\dots,T(m+1)$, as a random sample from pairs (Y_t,Y_{t-1}), $t=2,\dots,T$
  compute $A_{T,t}^{*b}(u)$, $B_{T,t}^{*b}(u)$ and $\widetilde{\delta}_{T,t}^{*b}(u)$ from (3.31), (3.30), and (3.32), respectively, for $t=1,\dots,T(m+1)$
  calculate the bootstrap test statistic $\widetilde{M}_{T,\gamma}^{*b}$
end for
estimate the critical value as the corresponding empirical quantile of $\widetilde{M}_{T,\gamma}^{*1},\dots,\widetilde{M}_{T,\gamma}^{*B}$

Output: critical value for the test for $\mathcal{H}_0^{INAR}$ from (3.16)

(C.4) Assume that there exists $0 \le \nu_0 < m$ such that $\{Y_t\}_{t\le T+t_0}$ is a stationary solution of (3.6) for $t_0 = \lfloor T\nu_0 \rfloor$ and $\{Y_{T+t_0+t}\}_{t\ge 1} =^D \{Y_t^0\}$, where $\{Y_t^0\}$ is a stationary ergodic sequence with $\mathrm{var}(Y_t^0) < \infty$.

This type of general alternative was considered by Kirch and Tadjuidje Kamgaing (2015), and it covers, as a particular case, model INAR(1) with a change in the distribution of ε_t and/or in the parameter α.

Theorem 3.6. *Let (C.2)–(C.4) hold and $\widehat{\Delta}_{T,t}$ be defined by (3.21) and $q_\gamma(\cdot)$, $\gamma \in [0,1/2)$ is from (1.13). Then for any $\nu_0 < s \le m$, and as $T\to\infty$,*

$$\frac{1}{T}\widehat{\Delta}_{T,\lfloor Ts\rfloor} \xrightarrow{P} \left(\frac{s-\nu_0}{s}\right)^2 \int_0^1 \left(g_{Y,0}(u_\alpha)g_Y^0(u) - g_{Y,0}(u)g_Y^0(u_\alpha)\right)^2 w(u)\mathrm{d}u, \tag{3.33}$$

where $g_{Y,0}(u)$ and $g_Y^0(u)$ denote the PGF of Y_t for $t \le T+t_0$, and $t > T+t_0$, respectively, and $u_\alpha = 1+\alpha(u-1)$.

The assertion implies consistency of the test which rejects for large values of $M_{T,\gamma}$ and also the validity of (1.16) for a large group of fixed alternatives where the right hand side of (3.33) is nonzero. For instance, the procedure is sensitive to a change in parameter α, or/and in the distribution of ε_t.

Let us now turn to the INARCH model in (3.23) and testing $\mathcal{H}_0^{PAR}$. For the behavior under $\mathcal{H}_0^{PAR}$ we pose the following conditions:

(D.1) Let $\{Y_t\}$ be the stationary solution of model (3.23) under $\mathcal{H}_0^{PAR}$.

(D.2) Let $r_0(y,\boldsymbol{\theta})$ in (3.24) be Lipschitz-continuous in y for all $\boldsymbol{\theta} \in A$ with the Lipschitz constant always strictly smaller than 1.

(D.3) Let $\widehat{\boldsymbol{\theta}}$ be an estimator of the parameter $\boldsymbol{\theta}$ based on training data $Y_1,\dots,Y_T$ satisfying $T\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^2 = O_P(1)$ as $T\to\infty$, where $\|\cdot\|$ denotes the Euclidean vector norm.

Theorem 3.7. *Let (D.1)–(D.3) and (C.3) hold and let $M_{T,\gamma}$ be defined as in (3.27) with $\widehat{\delta}_{T,t}(u)$ in (3.25) and $q_\gamma(\cdot)$ in (1.13) for $\gamma \in [0,1/2)$. Then*

$$M_{T,\gamma} \xrightarrow{D} \sup_{0<s<m/(m+1)} \frac{1}{s^{2\gamma}} \int_0^1 \widetilde{V}^2(s,u)w(u)\mathrm{d}u,$$

where $\left\{\widetilde{V}(s,t);\, s \in (0, m/(m+1)),\, u \in (0,1)\right\}$ is a zero-mean Gaussian process with the covariance structure

$$\mathsf{E}\left(\widetilde{V}(s_1,u_1), \widetilde{V}(s_2,u_2)\right) = \min(s_1,s_2)\tilde{\sigma}(u_1,u_2)$$

with

$$\tilde{\sigma}(u_1,u_2) = \mathsf{E}\Big[\Big(u_1^{Y_2} - \mathsf{E}(u_1^{Y_2}|Y_1)\Big)\Big(u_2^{Y_2} - \mathsf{E}(u_2^{Y_2}|Y_1)\Big)\Big].$$

The limit null distribution again depends on various unknown quantities, so one possibility for computation of $c_{\alpha,m}$ is to approximate the limit distribution by estimating the covariance kernel of the corresponding Gaussian limit process. However, we again recommend a bootstrap procedure which consistently estimates the asymptotic null distribution and which is easy to implement and less computationally intensive.

In the case of $\mathcal{H}_0^{PAR}$, the distribution of observations under the null hypothesis is known up to finitely many parameters, so it is straightforward to apply the parametric bootstrap resampling scheme as it was suggested by Fokianos and Fried (2010) and further investigated by Fokianos and Neumann (2013). The procedure is summarized in Algorithm 4. Under appropriate conditions, it may be shown that the conclusions of Theorem 3.5 also hold for the current setting.

Algorithm 4 Parametric bootstrap for a change detection in a Poisson INARCH model.

Input: Training data $Y_1, \ldots, Y_T$, observations $\{Y_k\}_{k=T+1}^{Tm}$. Tuning parameter $\gamma \in (0,1/2]$.
1: calculate the estimator $\widehat{\boldsymbol{\theta}}$ and the test statistic $M_{T,\gamma}$
2: **for** $b = 1$ to B **do**
3: generate sequentially pseudo–observations $Y_1^{*b}, \ldots, Y_{T(m+1)}^{*b}$, as $Y_1^{*b} = Y_1$ and
4: **for** $t = 2$ to $T(m+1)$ **do**
5: generate Y_t^{*b} from a Poisson distribution with mean $r_0(Y_{t-1}^{*b}, \widehat{\boldsymbol{\theta}})$
6: **end for**
7: calculate the value of the bootstrap test statistic $M_{T,\gamma}^{*b}$
8: **end for**
9: estimate the critical value as the corresponding empirical quantile of $M_{T,\gamma}^{*1}, \ldots, M_{T,\gamma}^{*B}$.
Output: critical value for the test for $\mathcal{H}_0^{PAR}$ from (3.24).

Finally, we consider the limit behavior of $M_{T,\gamma}$ under the following alternatives:

(D.4) Assume that there exists $0 \le \nu_0 < m$ such that

$\{Y_t\}_{t \le T+t_0}$ is a stationary solution of (3.23) without a change for $t_0 = \lfloor T\nu_0 \rfloor$ and $\{Y_{T+t_0+t}\}_{t\ge 0} =^D \{Y_t^0\}$, where $\{Y_t^0\}$ is a stationary ergodic sequence with $\mathrm{var}(Y_t^0) < \infty$.

Theorem 3.8. *Let (D.2)–(D.4) and (C3) hold and let $\widehat{\Delta}_{T,t}$ be defined by (3.26). Then*

$$\frac{1}{T}\widehat{\Delta}_{T,\lfloor Ts \rfloor} \xrightarrow{P} \left(\frac{s-\nu_0}{s}\right)^2 \int_0^1 \left(\mathsf{E}\Big[u^{Y_2^0} - \exp\{r_0(Y_1^0, \boldsymbol{\theta})(u-1)\}\Big]\right)^2 w(u)\mathrm{d}u, \tag{3.34}$$

for any $\nu_0 < s \leq m$*, as* $T \to \infty$*.*

The assertion implies consistency of the respective test and also (1.16) for a large group of fixed alternatives where the right hand side of (3.34) is nonzero.

Chapter 4

Multivariate time series modeled via copulas

Copulas provide a powerful tool for modeling multivariate distributions (Nelsen, 2006; Joe, 2014; Embrechts, 2009). The key idea behind copula models is to separate the modeling of marginal distributions from the modeling of dependence structure via the relation (1.7). Copula-based models have been considered for both uncopulasivariate and multivariate time series processes, see Patton (2009), Patton (2012) and Fan and Patton (2014) for some overviews. In this chapter, we focus on multivariate series with continuous observations and structural form (1.8). This approach has been successfully applied in recent years in risk management and in modeling the dependence among different economic and financial series, see Fan and Patton (2014).

Section 4.1 introduces the considered time series copula models that allow for the presence of external covariates, and describes basic approaches to the estimation of the individual components of the model. The estimation itself is considered in more detail n Section 4.2. For the copula function, both nonparametric and parametric approaches are considered. Section 4.3 summarizes some asymptotical results, which are essential for the construction of the goodness-of-fit test in Section 4.4. Section 4.5 contains some additional results on the maximum likelihood estimation of the copula parameter.

This chapter is based on results from Neumeyer et al. (2019) and Omelka et al. (2020). We refer to Nelsen (2006) for basic notation regarding copulas.

4.1 Copula time series models with exogenous covariates

Let $\{\boldsymbol{Y}_t\}$ be a k-dimensional time series of interest such that $\boldsymbol{Y}_t = (Y_{1t}, \ldots, Y_{kt})^\top$ which is complemented with q-dimensional vector series $\{\mathbf{Z}_t\}$ of exogenous covariates, $\mathbf{Z}_t = (Z_{1t}, \ldots, Z_{qt})^\top$. Let $\mathcal{F}_{t-1}$ be a σ-field of the past information generated by the past observations of both series, i.e., $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{Y}_k, \mathbf{Z}_k,\ k \leq t-1\}$. Assume that the first two conditional moments of $\boldsymbol{Y}_t$ given $\mathcal{F}_{t-1}$ are

$$\mathsf{E}[Y_{jt}|\mathcal{F}_{t-1}] = m_j(\boldsymbol{X}_t), \quad \mathsf{var}[Y_{jt}|\mathcal{F}_{t-1}] = \sigma_j^2(\boldsymbol{X}_t), \quad t \in \mathbb{Z},\ j = 1, \ldots, k,$$

where $\boldsymbol{X}_t$ is a d-dimensional $\mathcal{F}_{t-1}$ measurable random vector including possibly past values of the proces $\{\boldsymbol{Y}_t\}$ as well as past values of $\{\mathbf{Z}_t\}$, and the mean functions $m_j : \mathbb{R}^d \to \mathbb{R}$ as well as the volatility functions $\sigma_j : \mathbb{R}^d \to \mathbb{R}^+$ are unknown for all $j = 1,\ldots,k$.

The aim is to model the conditional distribution of $\boldsymbol{Y}_t$ given $\mathcal{F}_{t-1}$ using an AR-ARCH type model

$$Y_{jt} = m_j(\boldsymbol{X}_t) + \sigma_j(\boldsymbol{X}_t)\,\varepsilon_{ji},\ i = 1,\ldots,n,\ j = 1,\ldots,k, \tag{4.1}$$

where the innovations $\{\boldsymbol{\varepsilon}_t\}$ are assumed to be independent and identically distributed (iid) random vectors such that $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t},\ldots,\varepsilon_{kt})^\top$ satisfies

$$\mathsf{E}\varepsilon_{jt} = 0, \quad \mathsf{var}\,\varepsilon_{jt} = 1, \quad j = 1,\ldots,k,$$

and $\boldsymbol{\varepsilon}_t$ is independent of the past and present covariates $\boldsymbol{X}_k$, $k \le t$ for all t. The model can be rewritten in a vector form as

$$\boldsymbol{Y}_t = \boldsymbol{m}(\boldsymbol{X}_t) + \Sigma(\boldsymbol{X}_t)\boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{m}(\boldsymbol{x}) = \big(m_1(\boldsymbol{x}),\ldots,m_k(\boldsymbol{x})\big)^\top$ and $\Sigma(\boldsymbol{x}) = \mathrm{diag}\big\{\sigma_1(\boldsymbol{x}),\ldots,\sigma_k(\boldsymbol{x})\big\}$.

In what follows, we assume that $\boldsymbol{\varepsilon}_t$ has an absolutely continuous distribution on $\mathbb{R}^d$ with a joint cdf $F_{\boldsymbol{\varepsilon}}$. Due to Sklar's theorem there exists a unique copula $C : [0,1]^d \to [0,1]$ such that

$$F_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \mathsf{P}(\varepsilon_{1t} \le x_1,\ldots,\varepsilon_{kt} \le x_k) = C\big(F_{1\varepsilon}(x_1),\ldots,F_{k\varepsilon}(x_k)\big), \tag{4.2}$$

for all $\boldsymbol{x} = (x_1,\ldots,x_k)^\top \in \mathbb{R}^k$, where $F_{j\varepsilon}$ denotes the marginal cumulative distribution function (cdf) of ε_{jt}, $j = 1,\ldots,k$. It then follows from (4.1) that

$$\mathsf{P}(Y_{1t} \le y_1,\ldots,Y_{kt} \le y_k \,|\, \boldsymbol{X}_t = \boldsymbol{x}) = \mathsf{P}(\varepsilon_{1t} \le z_1,\ldots,\varepsilon_{kt} \le z_k) = C(F_{1\varepsilon}(z_1),\ldots,F_{k\varepsilon}(z_k)),$$

where

$$z_j = \frac{y_j - m_j(\boldsymbol{x})}{\sigma_j(\boldsymbol{x})}, \quad j = 1,\ldots,k.$$

The copula C reflects the conditional dependence structure of the k individual time series after removing influences of the conditional means and variances.

The joint conditional distribution of $\boldsymbol{Y}_t$ given $\boldsymbol{X}_t$ is completely specified by the following three components:

(i) the individual conditional mean and volatility functions m_j and σ_j, $j = 1,\ldots,k$,

(ii) the marginal distributions of the innovations $F_{j\varepsilon}$, $j = 1,\ldots,k$, and

(iii) the copula function.

The estimation of (i)–(iii) is typically conducted in three sequential steps either fully parametrically, or semiparametrically, or purely nonparametrically.

A *fully parametric approach* uses a parametric model for all threee components (i)–(iii). In that case, a natural estimation method is the maximum likelihood (ML), (Patton, 2006, 2012). However, such fully parametric approach is typically too restrictive in practice, so its use is rare.

The most common modeling approach is *semiparametric*, when the marginal conditional moments in (i) and the copula function in (iii) are modeled parametrically, while the marginals of the residuals in (ii) are estimated nonparametrically, usually using the empirical cumulative distribution function (ecdf) of the estimated residuals from the conditional models. Typical approach to (i) is to use ARMA-GARCH methodology. Such framework is referred to as Semiparametric COpula-based Multivariate DYnamic (SCOMDY), introduced in paper Chen and Fan (2006). Since its publication SCOMDY models have became very popular in particular in the econometric (time-series) literature, see, e.g., overviews in Trivedi et al. (2007); Patton (2012); Manner and Reznikova (2012); Rémillard (2017) or a recent generalization Nasri and Rémillard (2019).

A *fully nonparametric approach* then imposes no structures on any of the components (i)–(ii) and uses purely nonparametric methods for their estimation (Choroś et al., 2010).

We will treat the conditional mean functions m_j and volatility functions σ_j in step (i) nonparametrically. The estimation of the marginal cdfs in (ii) is conducted nonparametrically as well, whereas we take two different approaches to estimating the copula C in step (iii), namely a nonparametric approach, leading to a fully nonparametric estimation of the whole model, and a parametric approach, leading to a semiparametric estimation of the model. The comparison of the two estimators of the copula C will be used for a construction of a *goodness-of-fit test* for a specified parametric copula family.

Note that models (4.1) with nonparametric means m_j and volatilities σ_j are referred to as multivariate nonparametric CHARN (conditional heteroscedastic autoregressive nonlinear) models (Härdle et al., 1998), and they have been studied by various authors, see Fan and Yao (2005); Gao (2007) for overviews.

As the innovations are not observable, both copula estimators will be based on estimated residuals. We will show that the asymptotic distribution is not affected by the necessary pre-estimation of m_j and σ_j in (i). This remarkable result is intrinsic for copula estimation and it was already observed in (semi)parametric estimation of copula.

4.2 Model estimation

For the ease of presentation, we will focus on the case of two time series, i.e., $k = 2$. However, all results can be straightforwardly extended to general $k \geq 2$. The model (4.1) reduces to

$$Y_{1t} = m_1(\boldsymbol{X}_t) + \sigma_1(\boldsymbol{X}_t)\,\varepsilon_{1t}, \quad Y_{2t} = m_2(\boldsymbol{X}_t) + \sigma_2(\boldsymbol{X}_t)\,\varepsilon_{2t}. \tag{4.3}$$

Assume that the observed data $\left(\boldsymbol{Y}_t^\top, \boldsymbol{X}_t^\top\right)^\top$, $t = 1, \ldots, n$, come from a stationary stochastic process $\left\{(\boldsymbol{Y}_t^\top, \boldsymbol{X}_t^\top)^\top\right\}_{t\in\mathbb{Z}}$ so that (4.3) is satisfied for $t = 1, \ldots, n$. It follows from (4.2) that the copula function C is uniquely determined as

$$C(u_1, u_2) = F_\varepsilon\Big(F_{1\varepsilon}^{-1}(u_1), F_{2\varepsilon}^{-1}(u_2)\Big), \quad (u_1, u_2) \in [0,1]^2. \tag{4.4}$$

Since the innovations $\boldsymbol{\varepsilon}_t$ are unobserved, the inference about the copula function C is based on the estimated residuals $\widehat{\boldsymbol{\varepsilon}}_t = (\widehat{\varepsilon}_{1t}, \widehat{\varepsilon}_{2t})^\top$ with

$$\widehat{\varepsilon}_{jt} = \frac{Y_{jt} - \widehat{m}_j(\boldsymbol{X}_t)}{\widehat{\sigma}_j(\boldsymbol{X}_t)}, \qquad t = 1, \ldots, n, \quad j = 1, 2, \tag{4.5}$$

where $\widehat{m}_j$ and $\widehat{\sigma}_j$ are suitable estimates of the unknown functions m_j and σ_j.

In what follows we will use the *local polynomial estimators* of order P, see (Fan and Gijbels, 1996, Chapter 2.3) or Masry (1996). For a given $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$, $\widehat{m}_j(\boldsymbol{x})$ is defined as $\widehat{\beta}_{\mathbf{0}}$, the component of $\widehat{\boldsymbol{\beta}}$ with multi-index $\mathbf{0} = (0, \ldots, 0)$, where $\widehat{\boldsymbol{\beta}}$ is the solution to the minimization problem

$$\min_{\beta = (\beta_{\mathbf{i}})_{\mathbf{i} \in \mathbb{I}}} \sum_{t=1}^{n} \Big[Y_{jt} - \sum_{\mathbf{i} \in \mathbb{I}} \beta_{\mathbf{i}} \, \psi_{\mathbf{i}, \mathbf{h}_n}\big(\boldsymbol{X}_t - \boldsymbol{x}\big) \Big]^2 K_{\mathbf{h}_n}(\boldsymbol{X}_t - \boldsymbol{x}), \tag{4.6}$$

where $\mathbb{I} = \mathbb{I}(d, P)$ denotes the set of multi-indices $\mathbf{i} = (i_1, \ldots, i_d)$ with $i_1 + \cdots + i_d \leq P$ and

$$\psi_{\mathbf{i}, \mathbf{h}_n}(\boldsymbol{x}) = \prod_{k=1}^{d} \left(\frac{x_k}{h_n^{(k)}} \right)^{i_k} \frac{1}{i_k!},$$

and

$$K_{\mathbf{h}_n}(\boldsymbol{X}_t - \boldsymbol{x}) = \prod_{k=1}^{d} \frac{1}{h_n^{(k)}} \, k\left(\frac{X_{tk} - x_k}{h_n^{(k)}} \right),$$

with k being a kernel function and $\mathbf{h}_n = \left(h_n^{(1)}, \ldots, h_n^{(d)} \right)$ the smoothing parameter, see (Fan and Gijbels, 1996, Chapter 2.2) for more details on these notations.

The conditional variance $\sigma_j^2(\boldsymbol{x})$ is estimated as

$$\widehat{\sigma}_j^2(\boldsymbol{x}) = \widehat{s}_j(\boldsymbol{x}) - \widehat{m}_j^2(\boldsymbol{x}),$$

where $\widehat{s}_j(\boldsymbol{x})$ is obtained in the same way as $\widehat{m}_j(\boldsymbol{x})$ but with Y_{jt} replaced with Y_{jt}^2.

4.2.1 Nonparametric copula estimation

A natural nonparametric estimator of the copula function C is obtained if the true cdfs in (4.4) are replaced with suitable estimators. Since the true innovations ε_{jt} are not observed, these estimators have to be constructed from the estimated residuals. We propose to use slightly modified empirical cdfs.

Let $\{c_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers such that $c_n \to \infty$ as $n \to \infty$, and define $\mathbf{J}_n = [-c_n, c_n]^d$ and a weight function

$$w_n(\boldsymbol{x}) = \mathbf{1}\{\boldsymbol{x} \in \mathbf{J}_n\}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Furthermore, set

$$w_{nt} = w_n(\boldsymbol{X}_t), \quad t = 1, \ldots, n, \quad \text{and} \quad W_n = \sum_{t=1}^{n} w_{nt}.$$

Define

$$\widehat{F}_{\widehat{\varepsilon}}(y_1, y_2) = \frac{1}{W_n} \sum_{t=1}^{n} w_{nt}\, \mathbf{1}\Big\{\widehat{\varepsilon}_{1t} \leq y_1, \widehat{\varepsilon}_{2t} \leq y_2\Big\}, \tag{4.7}$$

and

$$\widehat{F}_{j\widehat{\varepsilon}}(y) = \frac{1}{W_n} \sum_{t=1}^{n} w_{nt}\, \mathbf{1}\Big\{\widehat{\varepsilon}_{jt} \leq y\Big\}, \quad j = 1, 2.$$

Then $\widehat{F}_{\widehat{\varepsilon}}$ and $\widehat{F}_{j\widehat{\varepsilon}}$ are nonparametric estimates of the joint distribution function F_{ε} and the marginal cdfs $F_{j\varepsilon}$, $j = 1, 2$, respectively. It is visible that these estimators behave asymptotically as the classical empirical cdfs computed from the estimated residuals. The weighting takes care of the observations with large values of the covariate vector $\boldsymbol{X}_t$, where the estimates $\widehat{m}_j(\boldsymbol{X}_t)$ and $\widehat{\sigma}_j(\boldsymbol{X}_t)$ are less reliable and so are the estimated residuals.

Finally, the resulting nonparametric estimator of C is

$$\widetilde{C}_n(u_1, u_2) = \widehat{F}_{\widehat{\varepsilon}}\Big(\widehat{F}_{1\widehat{\varepsilon}}^{-1}(u_1), \widehat{F}_{2\widehat{\varepsilon}}^{-1}(u_2)\Big), \quad (u_1, u_2)^{\top} \in (0, 1)^2. \tag{4.8}$$

4.2.2 Semiparametric copula estimation

Assume now that the copula function C belongs to a parametric family

$$\mathcal{C} = \Big\{C(\mathbf{u}; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\Big\}, \tag{4.9}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^{\top}$ is an unknown finite dimensional parameter and $\Theta \subset \mathbb{R}^p$. As noted in Section 4.1, a semiparametric estimation of the parameter $\boldsymbol{\theta}$ is popular within the copula setup, see Tsukahara (2005) for an iid setup, and Chen and Fan (2006) for SCOMDY models.

We consider estimation of $\boldsymbol{\theta}$ based on the method of moments (particularly on the inversion of Kendall's tau) and an M-estimation method, which involves the pseudo-maximum likelihood as a special case.

Method-of-Moments (inversion of Kendall's tau). Consider for simplicity that $p = 1$, i.e., the parameter θ is one-dimensional. The estimator $\widehat{\theta}_n^{(ik)}$ of θ based on the inversion of Kendall's tau is defined as

$$\widehat{\theta}_n^{(ik)} = \tau^{-1}(\widehat{\tau}_n),$$

where

$$\tau(\theta) = -1 + 4 \iint_{[0,1]^2} C(u_1, u_2; \theta)\, \mathrm{d}C(u_1, u_2; \theta)$$

is the theoretical Kendall's tau and $\widehat{\tau}_n$ is an (empirical) estimate of Kendall's tau computed from the estimated residuals $\widehat{\boldsymbol{\varepsilon}}_t$, see (Remillard, 2013, Chapter 8.5.3). Note that, in our setting, the Kendall's tau would be computed from the estimated residuals $(\widehat{\varepsilon}_{1t}, \widehat{\varepsilon}_{2t})$ for which $w_{nt} > 0$.

M-estimation. Denote as

$$\mathbf{U}_t^\top = \big(U_{1t}, U_{2t}\big) = \big(F_{1\varepsilon}(\varepsilon_{1t}), F_{2\varepsilon}(\varepsilon_{2t})\big). \tag{4.10}$$

and let

$$\widetilde{\mathbf{U}}_t^\top = \big(\widetilde{U}_{1t}, \widetilde{U}_{2t}\big) = \tfrac{W_n}{W_n+1}\Big(\widehat{F}_{1\widehat{\varepsilon}}\big(\widehat{\varepsilon}_{1t}\big), \widehat{F}_{2\widehat{\varepsilon}}\big(\widehat{\varepsilon}_{2t}\big)\Big) \tag{4.11}$$

be an observable counterpart of the unobserved random vector $\mathbf{U}_t$. Note that the multiplier $\frac{W_n}{W_n+1}$ is introduced in order to have both of the coordinates of the vector $\widetilde{\mathbf{U}}_t$ bounded away from zero and one.

The M-estimator of the parameter $\boldsymbol{\theta}$ is defined as

$$\widehat{\boldsymbol{\theta}}_n = \operatorname{argmin}_{\mathbf{t}\in\Theta} \sum_{t=1}^n w_{nt}\, \rho\big(\widetilde{\mathbf{U}}_t; \mathbf{t}\big)$$

where $\rho(\mathbf{u};\boldsymbol{\theta})$ is a given loss function. The estimator is usually computed as a solution to the estimating equations

$$\sum_{t=1}^n w_{nt}\, \boldsymbol{\phi}\big(\widetilde{\mathbf{U}}_t; \widehat{\boldsymbol{\theta}}_n\big) = \mathbf{0}, \tag{4.12}$$

where $\boldsymbol{\phi}(\mathbf{u};\boldsymbol{\theta}) = \partial\rho(\mathbf{u};\boldsymbol{\theta})/\partial\boldsymbol{\theta}$.

The class of M-estimators includes, among others, the *pseudo-maximum likelihood estimators* $\widehat{\boldsymbol{\theta}}_n^{(pl)}$, for which

$$\rho(\mathbf{u};\boldsymbol{\theta}) = -\log c(\mathbf{u};\boldsymbol{\theta}),$$

with $c(\cdot)$ being the copula density function, and so

$$\widehat{\boldsymbol{\theta}}_n^{(pl)} = \operatorname{argmax}_{\mathbf{t}\in\Theta} \sum_{t=1}^n w_{nt} \log c(\widetilde{\mathbf{U}}_t; \mathbf{t}).$$

4.3 Asymptotics

We first describe the asymptotic behavior of the nonparametric estimator $\widetilde{C}_n$ in Section 4.3.1. The asymptotics for the semiparametric estimation is then considered in Section 4.3.2.

Recall that it is assumed that $\boldsymbol{X}_t$ is d-dimensional for some $d \geq 1$. In this part we assume that

($\boldsymbol{\beta}$) The process $(\boldsymbol{X}_t, \boldsymbol{Y}_t)_{t\in\mathbb{Z}}$ is strictly stationary and absolutely regular (β-mixing) with the mixing coefficient β_n that satisfies $\beta_n = O(n^{-b})$ with $b > d+3$.

4.3.1 Asymptotics for the nonparametric copula estimator

The behavior of the nonparametric estimator $\widetilde{C}_n$ will be compared to the "oracle" estimator $C_n^{(or)}$ that would be computed from the unobserved innovations, i.e.

$$C_n^{(or)}(u_1, u_2) = \widehat{F}_\varepsilon\Big(\widehat{F}_{1\varepsilon}^{-1}(u_1), \widehat{F}_{2\varepsilon}^{-1}(u_2)\Big), \tag{4.13}$$

where

$$\widehat{F}_\varepsilon(z_1,z_2) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\left\{\varepsilon_{1t} \le z_1, \varepsilon_{2t} \le z_2\right\}$$

is the classical empirical estimator of F_ε based on the true unobserved innovations, and $\widehat{F}_{j\varepsilon}$ $(j = 1,2)$ are the corresponding marginal empirical cdfs.

For the formulation of the asymptotic properties of $\widetilde{C}_n$ from (4.8), we need to impose several technical regularity assumptions. An exact formulation of the assumptions, denoted as $(\mathbf{F}_\varepsilon)$, $(\mathbf{F}_\mathbf{X})$, $(\mathbf{Bw})$, $(\mathbf{M})$, $(\mathbf{k})$, $(\mathbf{J}_\mathbf{n})$ and $(\mathbf{m}\sigma)$, can be found in Neumeyer et al. (2019) which is attached to this thesis. A detailed discussion on these assumptions and some sufficient conditions for their validity is provided therein.

Theorem 4.1. *Suppose that assumptions* (β), $(\mathbf{F}_\varepsilon)$, $(\mathbf{F}_\mathbf{X})$, $(\mathbf{Bw})$, $(\mathbf{M})$, $(\mathbf{k})$, $(\mathbf{J}_\mathbf{n})$ *and* $(\mathbf{m}\sigma)$ *are satisfied. Then for* $n \to \infty$

$$\sup_{\mathbf{u}\in[0,1]^2} \sqrt{n}\left|\widetilde{C}_n(\mathbf{u}) - C_n^{(or)}(\mathbf{u})\right| \xrightarrow{P} 0.$$

The following Corollary follows immediately from Theorem 4.1 and the weak convergence of $\sqrt{n}\left[C_n^{(or)} - C\right]$ (see, e.g., Proposition 3.1 of Segers, 2012).

Corollary 4.1. *Under the assumptions of Theorem 4.1 the process*

$$\widetilde{\mathbb{C}}_n = \sqrt{n}\left[\widetilde{C}_n - C\right]$$

converges weakly in the space of bounded functions $\ell^\infty([0,1]^2)$ *to a centered Gaussian process* G_C*, which can be written as*

$$G_C(u_1,u_2) = B_C(u_1,u_2) - C^{(1)}(u_1,u_2)\,B_C(u_1,1) - C^{(2)}(u_1,u_2)\,B_C(1,u_2),$$

where B_C *is a Brownian bridge on* $[0,1]^2$ *with covariance function*

$$\mathsf{E}\left[B_C(u_1,u_2)B_C(u_1',u_2')\right] = C(u_1 \wedge u_1', u_2 \wedge u_2') - C(u_1,u_2)\,C(u_1',u_2'),$$

where $x \wedge y = \min\{x,y\}$ *and* $C^{(j)}$ *is the* j*-th firth order partial derivative of* C*,* $j = 1,2$.

4.3.2 Asymptotics for the semiparametric estimators

Section 4.2.2 defines semiparametric estimators of the copula parameter, specifically the Kendall's tau estimator $\widehat{\theta}_n^{(ik)}$ of a univariate parameter θ and an M-estimator $\widehat{\boldsymbol{\theta}}_n$ of a general p-dimensional parameter $\boldsymbol{\theta}$, with the pseudo-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n^{(pl)}$ being an important special case.

The next claim describes asymptotics for $\widehat{\theta}_n^{(ik)}$. The assertion follows immediately from Theorem 4.1 and Hadamard differentiability of Kendall's tau proved in Veraverbeke et al. (2011, Lemma 1).

Corollary 4.2. *Let the assumptions of Theorem 4.1 hold and let* C *belong to the parametric family (4.9). If* $\tau'(\theta) \neq 0$ *then* $\widehat{\theta}_n^{(ik)}$ *satisfies*

$$\sqrt{n}\left(\widehat{\theta}_n^{(ik)} - \theta\right) \xrightarrow{D} \mathcal{N}\left(0, \tfrac{\sigma_\tau^2}{[\tau'(\theta)]^2}\right),$$

where

$$\sigma_\tau^2 = \operatorname{var}\Big\{8\,C(U_{11},U_{21};\theta) - 4\,U_{11} - 4\,U_{21}\Big\},$$

and $\big(U_{11},U_{21}\big)^\top$ *is defined in* (4.10).

In order to describe asymptotics of an M-estimator $\widehat{\boldsymbol\theta}_n$, we first define its oracle counterpart computed from the true innovations. Let $\widehat{\boldsymbol\theta}_n^{(or)}$ be the consistent root of the estimating equations

$$\sum_{t=1}^{n} \phi\Big(\widehat{\mathbf{U}}_t;\widehat{\boldsymbol\theta}_n^{(or)}\Big) = \mathbf{0}, \tag{4.14}$$

where

$$\widehat{\mathbf{U}}_t^\top = \Big(\widehat{U}_{1t},\widehat{U}_{2t}\Big) = \tfrac{n}{n+1}\Big(\widehat{F}_{1\varepsilon}(\varepsilon_{1t}),\widehat{F}_{2\varepsilon}(\varepsilon_{2t})\Big) \tag{4.15}$$

are the standard pseudo-observations calculated from the unobserved innovations and their marginal empirical distribution functions $\widehat{F}_{1\varepsilon}(y)$ and $\widehat{F}_{2\varepsilon}(y)$.

We formulate general assumptions under which there exists a consistent root $\widehat{\boldsymbol\theta}_n$ of the estimating equations (4.12) that is asymptotically equivalent to $\widehat{\boldsymbol\theta}_n^{(or)}$. Since the special case of pseudo-maximum likelihood estimation is discussed in more detail within a slightly modified setup in Section 4.5, we provide the conditions explicitly here, see also Remark 4.1 below.

Let $\boldsymbol\theta$ stand for the true value of the parameter. Consider the following assumptions:

- **(Id)** $\boldsymbol\theta$ is a unique minimizer of the function $r(\mathbf{t}) = \mathsf{E}\,\rho(U_{1t},U_{2t};\mathbf{t})$ and $\boldsymbol\theta$ is an inner point of Θ.
- **($\boldsymbol\phi$)** There exists an open neighbourhood $V(\boldsymbol\theta)$ of $\boldsymbol\theta$ such that for each $l_1,l_2 \in \{1,\ldots,p\}$, the functions $\phi_{l_1}(u_1,u_2;\mathbf{t}) = \frac{\partial\rho(u_1,u_2;\mathbf{t})}{\partial t_{l_1}}$ and $\phi_{l_1,l_2}(u_1,u_2;\mathbf{t}) = \frac{\partial\rho(u_1,u_2;\mathbf{t})}{\partial t_{l_1}\partial t_{l_2}}$ are uniformly continuous in (u_1,u_2) uniformly in $\mathbf{t}\in V(\boldsymbol\theta)$ and of uniformly bounded Hardy-Kraus variation (see, e.g., Berghaus et al., 2017).
- **($\boldsymbol\phi^{(\mathrm{j})}$)** There exists $V(\boldsymbol\theta)$ and a function $h(u_1,u_2)$ such that for each $\mathbf{t}\in V(\boldsymbol\theta)$

  $$\max_{j=1,2}\max_{l=1,\ldots,p}\Big|\phi_l^{(j)}(u_1,u_2;\mathbf{t})\Big| \le h(u_1,u_2),\ \text{where } \phi_l^{(j)}(u_1,u_2;\mathbf{t}) = \tfrac{\partial\phi_l(u_1,u_2;\mathbf{t})}{\partial u_j}$$

  and $\mathsf{E}\,h(U_{11},U_{21}) < \infty$.
- **($\boldsymbol\Gamma$)** Each element of the (matrix) function $\boldsymbol\Gamma(\mathbf{t}) = \mathsf{E}\,\frac{\partial\boldsymbol\phi(U_1,U_2;\mathbf{t})}{\partial\mathbf{t}^\top}$ is a continuous function on $V(\boldsymbol\theta)$ and the matrix $\boldsymbol\Gamma = \boldsymbol\Gamma(\boldsymbol\theta)$ is positive definite.

Theorem 4.2. *Suppose that the assumptions of Theorem 4.1 are satisfied and that also* (**Id**), ($\boldsymbol\phi$), ($\boldsymbol\phi^{(\mathrm{j})}$), *and* ($\boldsymbol\Gamma$) *hold. Then with probability going to one as* $n\to\infty$ *there exists a consistent root* $\widehat{\boldsymbol\theta}_n$ *of the estimating equations* (4.12), *which satisfies*

$$\sqrt{n}\Big(\widehat{\boldsymbol\theta}_n - \boldsymbol\theta\Big) \xrightarrow{D} \mathcal{N}\Big(\mathbf{0},\boldsymbol\Gamma^{-1}\,\boldsymbol\Sigma\,\boldsymbol\Gamma^{-1}\Big), \tag{4.16}$$

where

$$\Sigma = \operatorname{var}\Big\{\phi\big(U_{11},U_{21};\boldsymbol{\theta}\big) + \iint \big[\mathbf{1}\{U_{11}\le v_1\} - v_1\big]\tfrac{\partial\phi(v_1,v_2;\boldsymbol{\theta})}{\partial v_1}\,dC(v_1,v_2;\boldsymbol{\theta})$$
$$+ \iint \big[\mathbf{1}\{U_{21}\le v_2\} - v_2\big]\tfrac{\partial\phi(v_1,v_2;\boldsymbol{\theta})}{\partial v_2}\,dC(v_1,v_2;\boldsymbol{\theta})\Big\}.$$

Note that the asymptotic distribution of the estimator $\widehat{\boldsymbol{\theta}}_n$ coincides with the distribution given in Section 4 of Genest et al. (1995) of the consistent root $\widehat{\boldsymbol{\theta}}_n^{(or)}$ of the estimating equations (4.14). Therefore, using the residuals instead of the true innovations has asymptotically negligible effect on the (first-order) asymptotic properties. In fact, it can be even shown that both $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\boldsymbol{\theta}}_n^{(or)}$ have the same asymptotic representations and thus

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n^{(or)}\big) = o_P(1).$$

Remark 4.1. Unfortunately, the assumptions of Theorem 4.2 exclude some useful models, e.g, the pseudo-maximum likelihood estimator in the Clayton family of copulas, for which the function $\boldsymbol{\phi}(\mathbf{u};\boldsymbol{\theta})$, viewed as a function of $\mathbf{u}$, is unbounded. This problem is further discussed in Section 4.5.

4.4 Goodness-of-fit testing

The semiparametric copula estimation from Section 4.2.2 requires a proper choice of the copula family $\mathcal{C}$ from (4.9). If the parametric family is not selected appropriately then the inference based on the semiparametric estimator can lead to incorrect conclusions. On the other hand, the nonparametric estimator is valid under general regularity assumptions. A comparison of the fully nonparametric estimator $\widetilde{C}_n$ with the semiparametric estimator $C(\cdot;\widehat{\boldsymbol{\theta}}_n)$ can be used for model verification and goodness-of-fit testing.

Let $\mathcal{C}_0 = \{C_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ be a given specified parametric family of copulas. The null hypothesis of interest can be formulated as

$$\mathcal{H}_0 : C \in \mathcal{C}_0 \tag{4.17}$$

against a general alternative $\mathcal{H}_1 : C \notin \mathcal{C}_0$. Various goodness-of-fit tests have been proposed for this problem in the iid setting (see, e.g., Genest et al., 2009; Kojadinovic and Holmes, 2009, and further references therein). The test statistic is typically constructed as a distance between two estimators of the copula function, a nonparametric one and a parametric one, leading to, for instance, a Kolmogorov-Smirnov test statistic or a Cramér von Mises test statistic, cf. (1.10) and (1.11).

Recall that we have shown that the asymptotic distribution of $\widetilde{C}_n(\mathbf{u})$ is the same as the asymptotic distribution of $\widetilde{C}_n^{(or)}(\mathbf{u})$, and the same holds for the estimator $\widehat{\boldsymbol{\theta}}_n$ and its oracle counterpart $\widehat{\boldsymbol{\theta}}_n^{(or)}$, where both estimators are either constructed using the method of moments or via the M-estimation. This allows us to use the results for iid data for the evaluation of (4.17). In particular, we propose to use the Cramér-von Mises statistic

$$S_n = \int_{[0,1]^2} \big[\widetilde{C}_n(\mathbf{u}) - C(\mathbf{u};\widehat{\boldsymbol{\theta}}_n)\big]^2\, d\widetilde{C}_n(\mathbf{u}), \tag{4.18}$$

and reject the null hypothesis $\mathcal{H}_0$ if

$$S_n > c_\alpha,$$

where the critical value c_α is determined using a parametric bootstrap so that $\mathsf{P}_{\mathcal{H}_0}(S_n > c_\alpha) \to \alpha$ as $n \to \infty$ for given $\alpha \in (0,1)$. The bootstrap proceeds as in the iid setup. It simply generates iid observations from the copula function $C(u_1, u_2; \widehat{\boldsymbol{\theta}}_n)$, and the test statistic is recalculated from these observations as if we directly observed the innovations. The procedure is summarized in Algorithm 5. It assumes that the tuning parameters for mean and volatility estimation (kernel k, bandwidth $\mathbf{h}_n$), and the sequence $\{c_n\}_{n=1}^{\infty}$ are given.

Note that, instead of generating n observations, we recommend to generate only W_n observations, because the copula is estimated, in fact, only from W_n observations and this should be reflected by the resampling procedure.

Algorithm 5 Parametric bootstrap for copula GOF.

Input: Data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ and $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.

1: compute the estimators $\widehat{m}_j, \widehat{\sigma}_j$ and the estimated residuals $\widehat{\varepsilon}_{jt}$, $j = 1, 2$, $t = 1, \ldots, n$
2: compute the nonparametric estimator $\widetilde{C}_n$
3: compute the estimator $\widehat{\boldsymbol{\theta}}_n$ under $\mathcal{H}_0$ and the test statistic S_n
4: **for** $b = 1$ to B **do**
5: generate $\mathbf{U}_t^{*b} = (U_{t1}^{*b}, U_{t2}^{*b})^\top$ for $t = 1, \ldots, W_n$ from distribution $C(u_1, u_2; \widehat{\boldsymbol{\theta}}_n)$
6: compute the estimator $\widehat{\theta}_n^{*b}$, $\widetilde{C}_n^{*b}$ and the test statistic S_n^{*b} from (4.18) from $\mathbf{U}_t^{*b}$
7: **end for**
8: cmpute the critical value c_α as the corresponding empirical quantile of S_n^{*b}, $b = 1, \ldots, B$, or calculate directly the p-value as

$$p = \frac{1}{B+1} \sum_{b=1}^{B} \mathrm{I}[S_n^{*b} \geq S_n]$$

Output: p-value of significance of $\mathcal{H}_0$ from (4.17)

Remark 4.2. A similar comparison of the nonparametric estimator $\widetilde{C}_n(\mathbf{u})$ with the independence copula $C(\mathbf{u}) = u_1 u_2$ provides a test for conditional independence of the two time series $\{Y_{1t}\}$ and $\{Y_{2t}\}$.

4.5 More on pseudo-maximum likelihood estimation

It is noted in Remark 4.1 that the assumptions of Theorem 4.2 are not fulfilled for the pseudo-maximum likelihood estimator in some important copula families where the function $\boldsymbol{\phi}(\mathbf{u}; \boldsymbol{\theta})$ is unbounded in $\mathbf{u}$. Hence, this section deals with an estimation in a simpler purely regression model and focuses on the formulation of appropriate regularity conditions that ensure validity of the claim formulated in Theorem 4.2. This Section is based on results from Omelka et al. (2020).

4.5.1 Considered setup

Assume now that the sequence $\{\boldsymbol{Y}_t\}$ is formed by independent k-dimensional random vectors and the d-dimensional vectors $\boldsymbol{X}_t$ in (4.1) are formed by purely exogenous covariates so that $(\boldsymbol{X}_t, \boldsymbol{Y}_t)$, $t = 1, \ldots, n$, can be assumed as independent and identically distributed (iid) pairs. Here, we focus on the situation when the margins follow parametric location-scale models

$$Y_{jt} = m_j(\boldsymbol{X}_t; \boldsymbol{\alpha}_j) + \sigma_j(\boldsymbol{X}_t; \boldsymbol{\alpha}_j)\,\varepsilon_{jt}, \quad j = 1, \ldots, k, \tag{4.19}$$

where $m_j(\boldsymbol{x}; \boldsymbol{\alpha}_j)$ and $\sigma_j(\boldsymbol{x}; \boldsymbol{\alpha}_j)$ are known functions depending only on an unknown (finite-dimensional) parameter $\boldsymbol{\alpha}_j$, $j = 1, \ldots, k$, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{kt})^\top$ is independent of $\boldsymbol{X}_t$ and satisfies the assumptions specified below (4.1). Namely, $\{\boldsymbol{\varepsilon}_t\}$ are iid with components having zero mean and unit variance such that the cumulative distribution function of $\boldsymbol{\varepsilon}_t$ is specified in (4.2). We further assume that the copula C belongs to a family of copulas

$$\mathcal{C} = \left\{C_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}$$

where $\boldsymbol{\Theta} \subset \mathbb{R}^p$. Our task is to estimate the true value of the copula parameter $\boldsymbol{\theta}$ based on the observations $\boldsymbol{X}_t, \boldsymbol{Y}_t$, $t = 1, \ldots, n$.

Since the errors $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ are not observed, the estimation is based on estimated residuals $\widehat{\boldsymbol{\varepsilon}}_t = (\widehat{\varepsilon}_{1t}, \ldots, \widehat{\varepsilon}_{kt})^\top$, $t = 1, \ldots, n$, where

$$\widehat{\varepsilon}_{jt} = \frac{Y_{jt} - m_j(\boldsymbol{X}_t; \widehat{\boldsymbol{\alpha}}_j)}{\sigma_j(\boldsymbol{X}_t; \widehat{\boldsymbol{\alpha}}_j)}, \quad t = 1, \ldots, n, \ j = 1, \ldots, k,$$

where $\widehat{\boldsymbol{\alpha}}_j$ is a suitable estimate of $\boldsymbol{\alpha}_j$.

The *maximum pseudo-likelihood estimator* of $\boldsymbol{\theta}$ based on the residuals is defined as

$$\widehat{\boldsymbol{\theta}}_n = \operatorname{argmax}_{\boldsymbol{t} \in \boldsymbol{\Theta}} \sum_{t=1}^{n} \log c\big(\widehat{\mathbf{U}}_t; \boldsymbol{t}\big),$$

where $c(\cdot; \boldsymbol{t})$ is the density of the copula $C_{\boldsymbol{t}} \in \mathcal{C}$, and

$$\widehat{\mathbf{U}}_t = \left(\widehat{U}_{1t}, \ldots, \widehat{U}_{kt}\right)^\top = \tfrac{n}{n+1}\left(\widehat{F}_{1\widehat{\varepsilon}}(\widehat{\varepsilon}_{1t}), \ldots, \widehat{F}_{k\widehat{\varepsilon}}(\widehat{\varepsilon}_{kt})\right)^\top. \tag{4.20}$$

As it is common in the maximum likelihood theory, we consider the estimator $\widehat{\boldsymbol{\theta}}_n$ to be an appropriately chosen root of the estimating equations

$$\sum_{t=1}^{n} \boldsymbol{\psi}(\widehat{\mathbf{U}}_t; \widehat{\boldsymbol{\theta}}_n) = \mathbf{0}, \quad \text{where} \quad \boldsymbol{\psi}(\mathbf{u}; \boldsymbol{t}) = \frac{\partial \log c\big(\mathbf{u}; \boldsymbol{t}\big)}{\partial \boldsymbol{t}}. \tag{4.21}$$

Analogously, the *oracle maximum pseudo-likelihood estimator* $\widetilde{\boldsymbol{\theta}}_n$ is based on the true (but unobserved) errors ε_{jt} and it is defined by the estimating equation (4.21) with $\widehat{\mathbf{U}}_t$ replaced by the true pseudo-observations

$$\widetilde{\mathbf{U}}_t = \left(\widetilde{U}_{1t}, \ldots, \widetilde{U}_{kt}\right)^\top = \tfrac{n}{n+1}\left(\widehat{F}_{1\varepsilon}(\varepsilon_{1t}), \ldots, \widehat{F}_{k\varepsilon}(\varepsilon_{kt})\right)^\top, \tag{4.22}$$

where $\widehat{F}_{j\varepsilon}$ is the marginal empirical distribution function of the (unobserved) errors, i.e.,

$$\widehat{F}_{j\varepsilon}(y) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{\varepsilon_{jt} \leq y\}, \quad j = 1,\ldots,k.$$

Even in this simplified setup, the derivation of the asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$ was surprisingly more technically involved than expected. The main reason is that the adaptation of the existing results is not at all straightforward as either the crucial steps in the proofs are, probably due to the broad scope of the presented results, missing (Chan et al., 2009) or because the presented assumptions prevent the major technical difficulties at the cost of avoiding common copula families, see, e.g., assumptions (C.2) and (C.3) of Kim et al. (2007) and the corresponding assumptions in Kim et al. (2008), or the assumption about the quadratic integrability in Theorem 2 in Nasri and Rémillard (2019). Therefore, the main result, formulated in Section 4.5.2, was derived with our own approach.

4.5.2 Asymptotics

The following two theorems state that, under specified regularity conditions, the estimator $\widehat{\boldsymbol{\theta}}_n$ based on the residuals has the same asymptotic distribution as the unattainable oracle estimator $\widetilde{\boldsymbol{\theta}}_n$. Thus when fitting the copula C one can (under the stated assumptions) ignore the fact that the true innovations were replaced with the estimated residuals.

First of all, we need to assume that the density of the error term ε_j is "well-behaving" on the border of its support.

(C.1). For each $j \in \{1,\ldots,k\}$ the density function $f_{j\varepsilon}$ of ε_j is continuous on the support of ε_j and there exists $\beta \in [0, \frac{1}{2})$ such that

$$\sup_{u\in(0,1)} \frac{f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(u)\big)\big(1+|F_{j\varepsilon}^{-1}(u)|\big)}{u^{\beta}(1-u)^{\beta}} < \infty \tag{4.23}$$

and

$$\sup_{u\in(0,1/2)} \frac{f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(2u)\big)}{f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(u)\big)} < \infty \quad \text{and} \quad \sup_{u\in(1/2,1)} \frac{f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(1-2u)\big)}{f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(1-u)\big)} < \infty.$$

Furthermore, for some u_1, u_2 in $(0,1)$, the function $f_{j\varepsilon}\big(F_{j\varepsilon}^{-1}(u)\big)$ is non-decreasing on $(0,u_1)$ and non-increasing on $(u_2,1)$.

Note that assumption **(C.1)** with $\beta = 0$ allows also for distributions with bounded discontinuous densities (e.g., exponential and uniform).

Remaining assumptions **(C.2)**–**(C.7)** are formulated in Section 4.5.4. Roughly speaking, assumption **(C.2)** states that the parametric models in (4.19) can be estimated at the standard $\sqrt{n}$-rate and that the functions m_j, σ_j are sufficiently smooth. Assumption **(C.3)** is a standard identification condition, while **(C.4)** requires smoothness of the function $\log[c(\mathbf{u},\boldsymbol{\theta})]$ in $\boldsymbol{\theta}$. Assumptions **(C.5)** and **(C.6)** are related to the behavior of the function $\boldsymbol{\psi}(\mathbf{u},\boldsymbol{t})$ and its derivative with respect to $\boldsymbol{t}$. Finally, **(C.7)** is a standard assumption on the Fisher information matrix.

Theorem 4.3. *Let (C.1) with $\beta > 0$ and (C.2)-(C.7) hold. Then, as $n \to \infty$, with probability going to one there exists a consistent root $\widehat{\boldsymbol{\theta}}_n$ of the estimating equations (4.21) such that*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma), \tag{4.24}$$

where

$$\Sigma = \boldsymbol{I}^{-1}(\boldsymbol{\theta})\,\mathsf{var}\left(\widetilde{\boldsymbol{\psi}}\big(\mathbf{U}, \boldsymbol{\theta}\big)\right)\boldsymbol{I}^{-1}(\boldsymbol{\theta}),$$

where $\boldsymbol{I}(\boldsymbol{\theta})$ is defined in (C.7) and $\widetilde{\boldsymbol{\psi}}(\mathbf{u}, \boldsymbol{\theta}) = \left(\widetilde{\psi}_1(\mathbf{u}, \boldsymbol{\theta}), \ldots, \widetilde{\psi}_p(\mathbf{u}, \boldsymbol{\theta})\right)^{\top}$, where

$$\widetilde{\psi}_k(\mathbf{u}, \boldsymbol{\theta}) = \psi_k(\mathbf{u}; \boldsymbol{\theta}) + \sum_{j=1}^{k} \int_{[0,1]^k} \left[\mathbb{1}\{u_j \le v_j\} - v_j\right] \psi_k^{(j)}(\mathbf{v}; \boldsymbol{\theta})\, \mathrm{d}C(\mathbf{v}, \boldsymbol{\theta}), \quad k = 1, \ldots, p, \tag{4.25}$$

and $\psi_k^{(j)}(\mathbf{u}, \boldsymbol{\theta}) = \partial \psi_k(\mathbf{u}, \boldsymbol{\theta}) / \partial u_j$, $j = 1, \ldots, k$.

The next theorem claims that if the assumptions **(C.5)** and **(C.6)** are replaced with a stronger assumption **(C.8)**, which requires the derivatives of ψ_k with respect to $\boldsymbol{\theta}$ to be bounded, then (4.24) stays valid also if the marginals satisfy **(C.1)** with $\beta = 0$. Thus, for instance, if one (rightly) assumes that C is a Frank copula then the marginal distributions of the errors are allowed to be also uniform or exponential.

Theorem 4.4. *Let (C.1)-(C.4), (C.7), (C.8) hold. Then the statement of Theorem 4.3 holds.*

A detailed discussion on the imposed regularity conditions is included in Omelka et al. (2020). Although the assumptions that guarantee (4.24) are mild, they are not satisfied for some combinations of commonly used copula functions and marginal densities of the errors. Roughly speaking, there exists an interplay between assumed the copula family and the assumptions on the marginal distributions of the errors. Under the stated assumptions an unbounded copula density has to be compensated with marginal densities that are well-behaved not only in the supports of the corresponding distributions, but also at the border points of the supports, as formulated in assumption **(C.1)**. Nevertheless, this assumption is considerably milder than, for instance, assumption A2 of Chan et al. (2009). Moreover, our assumptions allow the marginal densities to be even discontinuous at the border of the supports when combined with a copula with "a well-behaving density" (e.g., a Frank copula). Hence, our results complement the pioneering work of Chen and Fan (2006) with rigorous mathematical proofs and with weakening the necessary assumptions.

4.5.3 Practical results

The article Omelka et al. (2020) contains a simulation study that illustrates the theoretical conclusions and explores how the maximum pseudo-likelihood estimator behaves if the posed regularity assumptions are violated. The results reveal that, in agreement with the theoretical results, the maximum pseudo-likelihood estimator outperforms the moment estimator based on the inversion of Kendall's tau in situations for which our regularity assumptions are satisfied. On the other hand, the performance of the maximum pseudo-likelihood estimator may deteriorate significantly if the regularity assumptions are not met. The problems are generally worse for larger values of Kendall's tau (a stronger dependence). One should be particularly careful when fitting the Clayton copula or the Gumbel copula to data with non-regular margins (as exponential or uniform) combined with a strong dependence structure.

For such situations, we recommend to use a modified ML estimator, inspired by the estimator from Fermanian and Lopez (2018), computed as the maximum pseudo-likelihood estimator solely from $\widehat{\mathbf{U}}_t$ which lie in $[\delta_n, 1-\delta_n]^2$, where $\delta_n = Dn^{-1/\lambda}$. The choice $D = 1/4$ and $\lambda = 1.9$ is explored in Omelka et al. (2020) and the simulations indicate that this estimator performs better for Clayton or Gumbel copula with non-regular marginals. Therefore, it presents an interesting alternative to the "standard" pseudo maximum-likelihood estimator.

It is also interesting that even though the density of the Gaussian copula is unbounded, the PML estimator performs better than the moment estimator even when combined with non-regular marginals. This raises a question whether a milder assumption than **(C.1)** would be sufficient for the Gaussian copula.

4.5.4 Appendix: Regularity conditions for asymptotics of the pseudo-maximum likelihood estimator

This sections contains formulation of the regularity assumptions **(C.2)** –**(C.8)**.

Regularity assumptions on m_j and σ_j. Let $S_{\boldsymbol{X}} \subset \mathbb{R}^d$ be the support of $\boldsymbol{X}_t$.

(C.2). For each $j \in \{1,\ldots,k\}$ let the following conditions hold:

- $\widehat{\boldsymbol{\alpha}}_j$ is a $\sqrt{n}$-consistent estimate of the parameter $\boldsymbol{\alpha}_j$,
- functions $m_j(\boldsymbol{x};\mathbf{a})$ and $\sigma_j(\boldsymbol{x};\mathbf{a})$ are differentiable with respect to $\mathbf{a}$ and the derivatives are denoted as $m_j'(\boldsymbol{x};\mathbf{a})$ and $\sigma_j'(\boldsymbol{x};\mathbf{a})$,
- there exists a neighborhood $U(\boldsymbol{\alpha}_j)$ of the true value of the parameter $\boldsymbol{\alpha}_j$ such that

  $$\inf_{\boldsymbol{x}\in S_{\boldsymbol{X}},\mathbf{a}\in U(\boldsymbol{\alpha}_j)} \sigma_j(\boldsymbol{x};\mathbf{a}) > 0$$

  and there exists a function $M_j : S_{\boldsymbol{X}} \to \mathbb{R}$ such that for each $\boldsymbol{x} \in S_{\boldsymbol{X}}$

  $$\sup_{\mathbf{a}\in U(\boldsymbol{\alpha}_j)} \left\| \tfrac{m_j'(\boldsymbol{x};\mathbf{a})}{\sigma_j(\boldsymbol{x};\mathbf{a})} \right\| \leq M_j(\boldsymbol{x}), \qquad \sup_{\mathbf{a}\in U(\boldsymbol{\alpha}_j)} \left\| \tfrac{s_j'(\boldsymbol{x};\mathbf{a})}{\sigma_j(\boldsymbol{x};\mathbf{a})} \right\| \leq M_j(\boldsymbol{x}),$$

  and $\mathsf{E}\big[M_j(\boldsymbol{X})\big]^r < \infty$ for some $r \geq 2$,
- for each $K > 0$ the derivatives $m_j'(\boldsymbol{x};\mathbf{a})$ and $\sigma_j'(\boldsymbol{x};\mathbf{a})$, viewed as functions of $\mathbf{a}$, are continuous at $\boldsymbol{\alpha}_j$ uniformly in $\boldsymbol{x} \in \{\tilde{\boldsymbol{x}} \in S_{\boldsymbol{X}} : \|\tilde{\boldsymbol{x}}\| \leq K\}$.

Regularity assumptions on the copula family $\mathcal{C}$. To formulate the main regularity assumptions about the copula family, we introduce the following set of functions.

Definition (Class of $\mathcal{J}$- and $\widetilde{\mathcal{J}}^{\beta_1,\beta_2}$-functions). A function $\varphi : (0,1)^k \to \mathbb{R}$ is called a $\mathcal{J}$-*function* if φ is continuous on $(0,1)^k$ and there exist $\eta \in [0,1)$ and a finite constant M_1 such that for all $\mathbf{u} \in (0,1)^k$

$$|\varphi(u_1,\ldots,u_k)| \leq \sum_{j=1}^{k} \frac{M_1}{\left[\min\{u_j, 1-u_j\}\right]^{\eta}}.$$

Let $\beta_1 \in [0,1/2)$ and $\beta_2 \geq 0$ be fixed. We say that a function $\varphi : (0,1)^k \to \mathbb{R}$ is a $\widetilde{\mathcal{J}}^{\beta_1,\beta_2}$-*function* if it is continuous on $(0,1)^k$ and there exists a finite constant M_2 such that for all $\mathbf{u} \in (0,1)^d$

$$|\varphi(u_1,\ldots,u_d)| \leq \sum_{j=1}^{k} \frac{M_2}{\left[\min\{u_j, 1-u_j\}\right]^{\beta_1}},$$

and $\left|\varphi^{(j)}(u_1,\ldots,u_d)\right| u_j^{\beta_2}(1-u_j)^{\beta_2}$ is a $\mathcal{J}$-function for all $j \in \{1,\ldots,k\}$, where

$$\varphi^{(j)}(u_1,\ldots,u_k) = \frac{\partial \varphi(u_1,\ldots,u_k)}{\partial u_j}.$$

Recall that $\Theta \subset \mathbb{R}^p$, $\boldsymbol{\theta}$ is the true value of the parameter, and $c(\cdot;\boldsymbol{t})$ is a density corresponding to the copula function $C_{\boldsymbol{t}}$. Denote as $\psi_k(\mathbf{u};\boldsymbol{t})$ the kth element of the vector function

$$\boldsymbol{\psi}(\mathbf{u};\boldsymbol{t}) = \partial \log\{c(\mathbf{u};\boldsymbol{t})\}/\partial \boldsymbol{t}.$$

(C.3). $c(\mathbf{u};\boldsymbol{t}_1) = c(\mathbf{u};\boldsymbol{t}_2)$ for almost all $\mathbf{u} \in (0,1)^d$ only if $\boldsymbol{t}_1 = \boldsymbol{t}_2$.

(C.4). The function $\log\{c(\mathbf{u};\boldsymbol{t})\}$ is continuously differentiable with respect to $\boldsymbol{t}$ for all $\mathbf{u} \in (0,1)^d$.

(C.5). For each $k \in \{1,\ldots,p\}$, $\psi_k(\cdot;\boldsymbol{\theta})$ is a $\widetilde{\mathcal{J}}^{\beta_1,\beta_2}$-function, where $\beta > \max\{\beta_1 + \frac{1}{r-1}, \beta_2\}$, for β from (C.1) and r from (C.2).

(C.6). The function $\boldsymbol{\psi}(\mathbf{u};\boldsymbol{t})$ is assumed to be continuously differentiable with respect to $\boldsymbol{t}$ for all $\mathbf{u} \in (0,1)^k$. Further there exist an open neighborhood $\mathcal{U} \subset \boldsymbol{\Theta}$ of $\boldsymbol{\theta}$ and a dominating $\mathcal{J}$-function $h(\mathbf{u})$ such that $\partial \boldsymbol{\psi}(\mathbf{u};\boldsymbol{t})/\partial \boldsymbol{t}^\top$ is continuous in $(0,1)^k \times \mathcal{U}$ and

$$\max_{k,\ell \in \{1,\ldots,p\}} \sup_{\boldsymbol{t} \in \mathcal{U}} \left|\frac{\partial \psi_k(\mathbf{u};\boldsymbol{t})}{\partial t_\ell}\right| \leq h(\mathbf{u}).$$

(C.7). The $p \times p$ (Fisher information) matrix $\boldsymbol{I}(\boldsymbol{\theta}) = -\mathsf{E}\left\{\partial \boldsymbol{\psi}(\mathbf{U};\boldsymbol{t})/\partial \boldsymbol{t}^\top\big|_{\boldsymbol{t}=\boldsymbol{\theta}}\right\}$, where

$$\mathbf{U} = \left(U_1,\ldots,U_d\right)^\top = \left(F_{1\varepsilon}(\varepsilon_1),\ldots,F_{d\varepsilon}(\varepsilon_d)\right)^\top,$$

is finite and nonsingular.

(C.8). The function $\boldsymbol{\psi}(\mathbf{u};\boldsymbol{t})$ is bounded and continuously differentiable with respect to $\boldsymbol{t}$ for all $\mathbf{u} \in (0,1)^k$. Further there exists an open neighborhood $\mathcal{U}$ of $\boldsymbol{\theta}$ such that $\partial \boldsymbol{\psi}(\mathbf{u};\boldsymbol{t})/\partial \boldsymbol{t}^\top$ is continuous in $(0,1)^k \times \mathcal{U}$ and

$$\max_{k,\ell \in \{1,\ldots,p\}} \sup_{\boldsymbol{t} \in \mathcal{U}} \sup_{\mathbf{u} \in (0,1)^k} \left|\frac{\partial \psi_k(\mathbf{u};\boldsymbol{t})}{\partial t_\ell}\right| < \infty \quad \text{and} \quad \max_{j \in \{1,\ldots,d\}} \max_{k \in \{1,\ldots,p\}} \sup_{\mathbf{u} \in (0,1)^k} \left|\frac{\partial \psi_k(\mathbf{u};\boldsymbol{\theta})}{\partial u_j}\right| < \infty.$$

Bibliography

Ahmad, A. and Francq, C. (2016). Poisson QMLE of count time series models. *J. Time Series Anal.*, 37:291–314.

Al-Osh, M. and Alzaid, A. (1987). First–order integer–valued autoregressive (INAR(1)) process. *J. Time Series Anal.*, 8:261–275.

Aleksandrov, B. and Weiss, C. H. (2020). Testing the dispersion structure of count time series using Pearson residuals. *AStA Adv. Stat. Anal.*, 104:325–361.

Alzaid, A. and Al-Osh, M. (1988). First–order integer–valued autoregressive (INAR(1)) process: distributional and regression properties. *Stat. Neerl.*, 42:53–61.

Antoch, J., Gregoire, G., and Jarušková, D. (2004). Detection of structural changes in generalized linear models. *Statist. Probab. Lett.*, 69:315–332.

Aue, A. and Horváth, L. (2004). Delay time in sequential detection of change. *Statist. Probab. Lett.*, 267:221–231.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.*, 34(1):1–16.

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2006). Change-point monitoring in linear models with conditionally heteroscedastic errors. *Econom. J.*, 9:373–403.

Barczy, M., Ispány, M., Pap, G., Scotto, M., and Silva, M. (2010). Innovation outliers in INAR(1) models. *Comm. Statist. Theory Methods*, 39:3343–3362.

Baringhaus, L. and Henze, N. (1992). A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statist. Probab. Lett.*, 13(4):269–274.

Berghaus, B., Bücher, A., and Volgushev, S. (2017). Weak convergence of the empirical copula process with respect to weighted metrics. *Bernoulli*, 23(1):743–772.

Berkes, I., Gombay, E., Horváth, L., and Kokoszka, P. (2004). Sequential change-point detection in GARCH(p,q) models. *Econometric Theory*, 20(6):1140–1167.

Bierens, H. (1982). Consistent model specification tests. *J. Econometrics*, 20:105–134.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2016). *Time series analysis: forecasting and control*. John Wiley & Sons, 5th edition.

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer, 2nd edition.

Brockwell, P. J. and Davis, R. A. (2006). *Time series: theory and methods*. Springer.

Chan, N.-H., Chen, J., Chen, X., Fan, Y., and Peng, L. (2009). Statistical inference for multivariate residual copula of GARCH models. *Statist. Sinica*, 19:53–70.

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *J. Econometrics*, 135:125–154.

Choroś, B., Ibragimov, R., and Permiakova, E. (2010). Copula estimation. In *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, pages 77–91, Berlin. Springer.

Christou, V. and Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *J. Time Series Anal.*, 35(1):55–78.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, 64:1045–1065.

Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley, New York.

Davis, R., Huang, D., and Yao, Y. (1995). Testing for a change in the parameter values and order of an autoregressive model. *Ann. Statist.*, 23:282–304.

Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika*, 90(4):777–790.

Davis, R. A., Fokianos, K., Holan, S. H., Joe, H., Livsey, J., Lund, R., Pipiras, V., and Ravishanker, N. (2021). Count time series: A methodological review. *J. Amer. Statist. Assoc.*, 116(535):1533–1547.

Davis, R. A., Holan, S. H., Lund, R., and Ravishanker, N. (2015). *Handbook of discrete-valued time series*. Chapman and Hall/CRC, New York.

Davis, R. A. and Liu, H. (2016). Theory and inference for a class of observation- driven models with application to time series of counts. *Statist. Sinica*, 26(4):1673–1707.

de Jong, R. M. and Woutersen, T. (2011). Dynamic time series binary choice. *Econometric Theory*, 27:673–702.

Doukhan, P. and Kengne, W. (2015). Inference and testing for structural change in general Poisson autoregressive models. *Electron. J. Stat.*, 9:1267–1314.

Du, J. G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *J. Time Series Anal.*, 12:129–142.

Embrechts, P. (2009). Copulas: A personal view. *J. Risk Insur.*, 76(3):639–650.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall/CRC, London.

Fan, J. and Yao, Q. (2005). *Nonlinear time series: nonparametric and parametric methods*. Springer series in statistics. Springer, New York.

Fan, Y. and Patton, A. J. (2014). Copulas in econometrics. *Annu. Rev. Econ.*, 6:179–200.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer–valued GARCH process. *J. Time Series Anal.*, 27:923–942.

Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas. *J. Multivariate Anal.*, 95(1):119–152.

Fermanian, J.-D. (2013). An overview of the goodness-of-fit test problem for copulas. In *Copulae in Mathematical and Quantitative Finance: Proceedings of the Workshop Held in Cracow, 10-11 July 2012*, pages 61–89. Springer.

Fermanian, J.-D. and Lopez, O. (2018). Single-index copulas. *J. Multivariate Anal.*, 165:27–55.

Fokianos, K. (2012). Count time series. In Subba Rao, T., Subba Rao, S., and Rao, C. R., editors, *Handbook of statistics 30: Time series – methods and applications*, pages 315–347. Elsevier, Amsterdam.

Fokianos, K. (2015). Statistical analysis of count time series models: A glm perspective. In Davis, R. A., Holan, S. H., Lund, R., and Ravishanker, N., editors, *Handbook of discrete-valued time series*, pages 3–28. Chapman and Hall/CRC.

Fokianos, K. and Fried, R. (2010). Interventions in INGARCH processes. *J. Time Series Anal.*, 31:210–225.

Fokianos, K. and Neumann, M. (2013). A goodness–of–fit test for Poisson count processes. *Electron. J. Stat.*, 7:793–819.

Fokianos, K. and Tjøstheim, D. (2009). Poisson autoregression. *J. Amer. Statist. Assoc.*, 104:1430–1439.

Fokianos, K. and Tjøstheim, D. (2011). Log-linear Poisson autoregression. *J. Multivariate Anal.*, 102(3):563–578.

Fokianos, K. and Tjøstheim, D. (2012). Nonlinear Poisson autoregression. *Ann. Inst. Statist. Math.*, 64:1205–1225.

Francq, C. and Zakoian, J.-M. (2019). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.

Franke, J., Kirch, C., and Kamgaing, J. T. (2012). Changepoints in times series of counts. *J. Time Series Anal.*, 33(5):757–770.

Freeland, R. (1998). *Statistical Analysis of Discrete Time series with Application to the Analysis of Workers' Compensation Claims Data*. PhD thesis, Management Science Division, Faculty of Commerce and Business Administration, University of British Columbia.

Gao, J. (2007). *Nonlinear time series: semiparametric and nonparametric methods*. Chapman & Hall/CRC, Boca Raton.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552.

Genest, C. and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Statist.*, 33:337–366.

Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness–of–fit testing in semiparametric models. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44:1096–1127.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance Math. Econom.*, 44(2):199–213.

Gürtler, N. and Henze, N. (2000). Recent and classical goodness-of-fitt tests for the Poisson distribution. *J. Statist. Plann. Inference*, 90:207–225.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Härdle, W., Tsybakov, A., and Yang, L. (1998). Nonparametric vector autoregression. *J. Statist. Plann. Inference*, 68(2):221–245.

Hlávka, Z., Hušková, M., Kirch, C., and Meintanis, S. (2016). Bootstrap procedures for online monitoring of changes in autoregressive models. *Comm. Statist. Simulation Comput.*, 45(7):2471–2490.

Horváth, L. (1989). The limit distributions of likelihood ratio and cumulative sum tests for a change in a binomial probability. *J. Multivariate Anal.*, 31(1):148–159.

Horváth, L. (1993). Change in autoregressive processes. *Stoch. Process. Appl.*, 44:221–242.

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitoring changes in linear models. *J. Statist. Plann. Inference*, 126:225–251.

Hudecová, Š. (2013a). Binary dynamic response model with structural changes. In *Proceedings of the 15th Applied Stochastic Models and Data Analysis (ASMDA 2013)*, pages 465–471. ASMDA International.

Hudecová, Š. (2013b). Structural changes in autoregressive models for binary time series. *J. Statist. Plann. Inference*, 143(10):1744–1752.

Hudecová, Š., Hušková, M., and Meintanis, S. (2015). Detection of changes in INAR models. In Steland, A., Rafajłowicz, E., and Szajowski, K., editors, *Stochastic Models, Statistics and their Applications*, volume 122 of *Springer Proceedings in Mathematics & Statistics*, pages 11–18, New York. Springer.

Hudecová, Š., Hušková, M., and Meintanis, S. (2015). Tests for time series of counts based on the probability-generating function. *Statistics*, 49(2):316–337.

Hudecová, Š., Hušková, M., and Meintanis, S. (2016). Change detection in INARCH time series of counts. In Cao, R., Manteiga, W. G., and Romo, J., editors, *Nonparametric Statistics: 2nd ISNPS, Cádiz, June 2014*, volume 175, pages 47–58, New York. Springer.

Hudecová, Š., Hušková, M., and Meintanis, S. (2017). Tests for structural changes in time series of counts. *Scand. J. Statist.*, 44(4):843–865.

Hudecová, Š., Hušková, M., and Meintanis, S. (2021). Goodness-of-fit tests for bivariate time series of counts. *Econometrics*, 9(1):10.

Hudecová, Š., Hušková, M., and Meintanis, S. G. (2024). Specifications tests for count time series models with covariates. *TEST*. https://doi.org/10.1007/s11749-024-00933-x.

Hušková, M. and Koubková, A. (2005). Monitoring jump changes in linear models. *J. Statist. Res.*, 39(2):51–70.

Hušková, M. and Meintanis, S. (2006). Change point analysis based on empirical characteristic functions of ranks. *Sequential Anal.*, 25:421–436.

Hušková, M., Prášková, Z., and Steinebach, J. (2007). On the detection of changes in autogeressive time series I. Asymptotics. *J. Statist. Plann. Inference*, 137:1243–1259.

Joe, H. (2014). *Dependence modeling with copulas*. CRC press.

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate discrete distributions*. John Wiley & Sons, Hoboken, New Jersey, 3rd edition.

Jung, R. and Tremayne, A. (2011). Useful models for time series or simply wrong ones? *AStA Adv. Stat. Anal.*, 95(1):59–91.

Jung, R. C., Kukuk, M., and Liesenfeld, R. (2006). Time series of count data: modeling, estimation and diagnostics. *Comput. Statist. Data Anal.*, 51(4):2350–2364.

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.

Kang, J. and Lee, S. (2014). Parameter change test for Poisson autoregressive models. *Scand. J. Statist.*, 41:1136–1152.

Kauppi, H. and Saikkonen, P. (2008). Predicting U.S. recessions with dynamic binary response models. *Rev. Econ. Stat.*, 90(4):777–791.

Kedem, B. and Fokianos, K. (2002). *Regression models for time series analysis*. Wiley, New York.

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007). Semiparametric estimation of the error distribution in multivariate regression using copulas. *Aust. N. Z. J. Stat.*, 49(3):321–336.

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2008). Estimating the error distribution in multivariate heteroscedastic time-series models. *J. Statist. Plann. Inference*, 138(5):1442–1458.

Kirch, C. and Tadjuidje Kamgaing, J. (2015). Detection of change points in discrete valued time series. In Davis, R. A., Holan, S. H., Lund, R., and Ravishanker, N., editors, *Handbook of Discrete Valued Time Series*, pages 219–244. Chapman and Hall/CRC.

Kocherlakota, S. and Kocherlakota, K. (1986). Goodness of fit tests for discrete distributions. *Comm. Statist. Theory Methods*, 15(3):815–829.

Kojadinovic, I. and Holmes, M. (2009). Tests of independence among continuous random vectors based on Cramér–von Mises functionals of the empirical copula process. *J. Multivariate Anal.*, 100(6):1137–1154.

Kokoszka, P. and Teyssière, G. (2002). Change-point detection in GARCH models: asymptotic and bootstrap tests. Available online at http://ideas.repec.org/p/cor/louvco/2002065.html.

Lee, S. and Kim, B. (2021). Recent progress in parameter change test for integer-valued time series models. *J. Korean Statist. Soc.*, 50:730–755.

Lehmann, E. L., Romano, J. P., and Casella, G. (2005). *Testing statistical hypotheses*. Springer, 3rd edition.

Leisch, F., Hornik, K., and Kuan, C.-M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econ. Theory*, 16:835–854.

Leucht, A. and Neumann, M. (2013). Degenerate U- and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. *Ann. Inst. Statist. Math.*, 65:349–386.

Ling, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *Ann. Statist.*, 35(3):1213–1237.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Ma, L. (1997). The asymptotic distributions of maximum likelihood ratio test and maximally selected χ^2-test in binomial observations. *J. Statist. Plann. Inference*, 67(1):17–43.

Manner, H. and Reznikova, O. (2012). A survey on time-varying copulas: Specification, simulations, and application. *Econometric Rev.*, 31(6):654–687.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Series Anal.*, 17(6):571–599.

McCullagh, P. and Nelder, J. (1989). *Generalized linear models*. Chapman and Hall, 2nd edition.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resour. Bull.*, 21:645–650.

McKenzie, E. (2003). Discrete variate time series. *Handbook of Statistics*, 21:573–606.

Meintanis, S. G. and Bassiakos, Y. (2005). Goodness-of-fit tests for additively closed count models with an application to the generalized Hermite distribution. *Sankhya*, 67:538–552.

Meintanis, S. G. and Nikitin, Y. Y. (2008). A class of count models and a new test for the Poisson distribution consistent in it. *J. Statist. Plann. Inference*, 138:3722–3732.

Nakamura, M. and Pérez-Abreu, V. (1993). Empirical probability generating function: An overview. *Insurance Math. Econom.*, 12:287–295.

Nakamura, M. and Pérez-Abreu, V. (1993). Use of an empirical probability generating function for testing a Poisson model. *Canad. J. Statist.*, 21(2):149–156.

Nasri, B. R. and Rémillard, B. N. (2019). Copula-based dynamic models for multivariate time series. *J. Multivariate Anal.*, 172:107–121.

Nelsen, R. B. (2006). *An Introduction to Copulas.* Springer, New York, 2nd edition.

Neumann, M. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli*, 17:1268–1284.

Neumeyer, N., Omelka, M., and Hudecová, Š. (2019). A copula approach for dependence modeling in multivariate nonparametric time series. *J. Multivariate Anal.*, 171:139–162.

Novoa-Muñoz, F. and Jiménez-Gamero, M. D. (2014). Testing for the bivariate Poisson distributions. *Metrika*, 77:771–793.

Novoa-Muñoz, F. and Jiménez-Gamero, M. D. (2016). A goodness-of-fit test for the multivariate Poisson distribution. *SORT*, 40:113–138.

Omelka, M., Gijbels, I., and Veraverbeke, N. (2009). Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Ann. Statist.*, 37:3023–3058.

Omelka, M., Hudecová, Š., and Neumeyer, N. (2020). Maximum pseudo-likelihood estimation based on estimated residuals in copula semiparametric models. *Scand. J. Statist.*, 48:1433–1473.

Patton, A. J. (2006). Estimation of multivariate models for time series of possibly different lengths. *J. Appl. Econometrics*, 21(2):147–173.

Patton, A. J. (2009). Copula-based models for financial time series. *Handbook of Financial Time Series*, 2009:767–785.

Patton, A. J. (2012). A review of copula models for economic time series. *J. Multivariate Anal.*, 110:4–18.

Pavlopoulos, H. and Karlis, D. (2008). INAR(1) modeling of overdispersed count series with an environmental application. *Environmetrics*, 19:369–393.

Pettitt, A. N. (1980). A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika*, 67(1):79–84.

Remillard, B. (2013). *Statistical methods for financial engineering*. CRC press.

Rémillard, B. (2017). Goodness-of-fit tests for copulas of multivariate time series. *Econometrics*, 5(1).

Rémillard, B., Papageorgiou, N., and Soustra, F. (2012). Copula-based semiparametric models for multivariate time series. *J. Multivariate Anal.*, 110:30–42.

Rémillard, B. and Theodorescu, R. (2000). Inference based on the empirical probability generating function for mixtures of Poisson distributions. *Stat. Risk Model.*, 18:349–366.

Schweer, S. and Weiss, C. (2014). Compound poisson inar(1) processes: Stochastic properties and testing for overdispersion. *Comput. Statist. Data Anal.*, 77:267–284.

Scotto, M. G., Weiss, C. H., and Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Stat. Model.*, 15:590–618.

Segers, J. (2012). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, 18:764–782.

Serbinowska, M. (1996). Consistency of an estimator of the number of changes in binomial observations. *Statist. Probab. Lett.*, 29:337–344.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Annales de l'ISUP*, 8(3):229–231.

Startz, R. (2008). Binomial autoregressive moving average models with an application to U.S. recession. *J. Bus. Econom. Statist.*, 26:1–8.

Szabó, T. (2011). Test statistics for parameter changes in INAR(p) models and a simulation study. *Austrian J. Stat.*, 40(4):265–280.

Trivedi, P. K., Zimmer, D. M., et al. (2007). *Copula modeling: an introduction for practitioners*. Now Publishers, Hanover, MA.

Tsay, R. S. (2002). *Analysis of financial time series*. John Wiley & Sons.

Tsay, R. S. (2013). *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.*, 33:357–375.

Veraverbeke, N., Omelka, M., and Gijbels, I. (2011). Estimation of a conditional copula and association measures. *Scand. J. Statist.*, 38:766–780.

Wang, C. and Li, W. K. (2011). On the autopersistence functions and the autopersistence graphs of binary autoregressive time series. *J. Time Series Anal.*, 32(6):639–646.

Weiss, C. (2011). Detecting mean changes in Poisson INAR(1) processes with EWMA control charts. *J. Appl. Statist. Sci.*, 38:383–398.

Weiss, C. and Testik, M. (2009a). CUSUM monitoring of first–order integer–valued autoregressive processes of Poisson counts. *J. Qualit. Technol.*, 41:389–400.

Weiss, C. and Testik, M. (2009b). CUSUM monitoring of first-order integer-valued autoregressive processes of Poisson counts. *J. Quality Techn.*, 41:389–400.

Weiss, C. and Testik, M. (2011). The Poisson INAR(1) CUSUM chart under overdispersion and estimation error. *IIE Transactions*, 43:805–818.

Weiss, C. and Testik, M. (2012). Detection of abrupt changes in count data time series: Cumulative sum derivations for INARCH(1) models. *J. Qualit. Technol.*, 44:249–264.

Weiss, C. and Testik, M. (2015). On the Phase I analysis for monitoring time–dependent count processes. *IIE Transactions*, 47:294–306.

Weiss, C. H. (2018a). *Discrete-valued time series*. John Wiley & Sons.

Weiss, C. H. (2018b). Goodness–of–fit testing of a count series' marginal distribution. *Metrika*, 81:619–651.

Weiss, C. H., Homburg, A., and Puig, P. (2019). Testing for zero inflation and overdispersion in INAR(1) models. *Statist. Papers*, 60:473–498.

Weiss, C. H. and Schweer, S. (2015). Detecting overdispersion in INARCH(1) processes. *Stat. Neerl.*, 69:281–297.

Wilks, D. and Wilby, R. (1999). The weather generation game: a review of stochastic weather models. *Prog. Phys. Geog.*, 23(3):329–357.

Worsley, K. J. (1983). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. *Biometrika*, 70(2):455–464.

Zeger, S. and Quaqish, B. (1988). Markov regression models for time series: a quasilikelihood approach. *Biometrics*, 44:1019–1031.

Zeileis, A. (2005). A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals. *Econometric Rev.*, 24:445–466.

Zhu, F. (2011). A negative binomial integer–valued GARCH model. *J. Time Series Anal.*, 32:54–67.

Appendix: Attached papers

HUDECOVÁ, Š. (2013). Structural changes in autoregressive models for binary time series. *Journal of Statistical Planning and Inference*, **143**(10), 1744–1752, doi:10.1016/j.jspi.2013.05.009.

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, doi:10.1080/02331888.2014.979826.

HUDECOVÁ, Š., HUŠKOVÁ, M. AND MEINTANIS, S. G. (2017). Tests for structural changes in time series of counts. *Scandinavian Journal of Statistics*, **44**(4), 843–865, doi:10.1111/sjos.12278.

NEUMEYER, N., OMELKA, M., AND HUDECOVÁ, Š. (2019). A copula approach for dependence modeling in multivariate nonparametric time series. *Journal of Multivariate Analysis*, **171**, 139–162, doi:10.1016/j.jmva.2018.11.016

OMELKA, M., HUDECOVÁ, Š., AND NEUMEYER, N. (2020). Maximum pseudo-likelihood estimation based on estimated residuals in copula semiparametric models. *Scandinavian Journal of Statistics*, **48**, 1433–1473, doi:10.1111/sjos.12498.