

Article

# Goodness-of-Fit Tests for Bivariate Time Series of Counts

Šárka Hudecová <sup>1,\*</sup> , Marie Hušková <sup>1,\*</sup>  and Simos G. Meintanis <sup>2,3</sup>

<sup>1</sup> Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Prague 8, Czech Republic

<sup>2</sup> Department of Economics, National and Kapodistrian University of Athens, 105 59 Athens, Greece; simosmei@econ.uoa.gr

<sup>3</sup> Unit for Pure and Applied Analytics, North-West University, Potchefstroom 2531, South Africa

\* Correspondence: hudecova@karlin.mff.cuni.cz (Š.H.); huskova@karlin.mff.cuni.cz (M.H.)

**Abstract:** This article considers goodness-of-fit tests for bivariate INAR and bivariate Poisson autoregression models. The test statistics are based on an L2-type distance between two estimators of the probability generating function of the observations: one being entirely nonparametric and the second one being semiparametric computed under the corresponding null hypothesis. The asymptotic distribution of the proposed tests statistics both under the null hypotheses as well as under alternatives is derived and consistency is proved. The case of testing bivariate generalized Poisson autoregression and extension of the methods to dimension higher than two are also discussed. The finite-sample performance of a parametric bootstrap version of the tests is illustrated via a series of Monte Carlo experiments. The article concludes with applications on real data sets and discussion.

**Keywords:** goodness-of-fit; bivariate time series; probability generating function; time series of counts



**Citation:** Hudecová, Šárka, Marie Hušková, and Simos G. Meintanis. 2021. Goodness-of-Fit Tests for Bivariate Time Series of Counts. *Econometrics* 9: 10. <https://doi.org/10.3390/econometrics9010010>

Received: 4 December 2020  
Accepted: 26 February 2021  
Published: 4 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Time series of counts enjoy numerous applications in such diverse fields as business, economics, engineering, and epidemiology, and corresponding inferential procedures have been intensively studied in recent years. The reader is referred to the earlier work of McKenzie (2003), as well as to the much updated full-book treatment of Davis et al. (2015) and Weiss (2018a) for an overview of the subject. Among the most popular models for count-time series are the integer autoregression (INAR) model and the Poisson autoregressive (PAR) model, also termed Poisson INARCH model. The INAR and PAR models, originally conceived for univariate counts, have been extended in order to accommodate bivariate, and more generally multivariate counts; see Latour (1997) and Liu (2012), respectively. One of the basic parametric elements of these generalizations involves moving from a univariate to a bivariate distribution, and eventually to a multivariate one.

In this connection, when confronted with a vectorial data-set of time series of counts, and since one is presented with a choice of several possible candidate models, it is extremely important to check model-adequacy via some goodness-of-fit (GOF) test. Otherwise a poorly fitted model might yield misleading inference. Various such procedures have been proposed in the literature for univariate series of counts. A brief overview of these approaches is provided in Section 2.

Inspired by the univariate test criteria, we propose GOF tests for bivariate INAR and PAR models. Our tests mainly target the (by far most popular) Poisson specification of these models, but also involve other structural aspects, such as order and functional specification of the underlying model. The suggested test statistic is constructed as a contrast between an estimator of the probability generating function (PGF) which is entirely model-free and a semiparametric counterpart that “respects” the model under test. Hence each new test may be viewed as a bivariate extension of the earlier PGF-based procedure suggested by Meintanis and Karlis (2014), Hudecová et al. (2015) and Schweer (2016).

The remainder of this work unfolds as follows. Section 2 provides a review of the models considered and of goodness-of-fit testing for univariate time series of counts. In Sections 3 and 4 we introduce the bivariate models and the new test criteria. In Section 5 the asymptotic properties of the methods are investigated and resampling versions of the tests are proposed. An extension of one the tests to a more complicated model is discussed in Section 6. The finite-sample properties of the new criteria are studied by means of Monte Carlo methods in Section 7. Applications to real data sets are included in Section 8 while our final thoughts regarding the news methods and potential extensions thereof are summarized in Section 9. Proofs of asymptotics are provided in Appendix A.

## 2. Goodness-of-Fit Methods for Univariate Time Series of Counts

Many of the models for time series of low counts assume that conditionally on the past, the distribution of the current variable is fully specified by a family of laws indexed by a certain parameter. An important example is a class of integer autoregressive conditionally heteroscedastic (INGARCH) models, covering PAR models as special cases. Different models are based on the thinning operator see [Steutel and van Harn \(1979\)](#), with the INAR model being the most popular one. 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.

Let  $\{Y_t\} = \{Y_t, t \in \mathbb{Z}\}$  be a univariate time series and let  $\mathcal{F}_t = \sigma\{Y_s, s \leq t\}$  is the information known up to time  $t$ . An integer GARCH model, abbreviated as nonlinear INGARCH( $p, q$ ), is defined as

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

where  $F(\lambda)$  is some count distribution with mean  $\lambda$ , and the regression function  $r : [0, \infty)^{p+q} \times \Theta$  belongs to some specific parametric family of functions  $\mathcal{R} = \{r(\cdot, \boldsymbol{\eta}) : \boldsymbol{\eta} \in \Theta \subset \mathbb{R}^k\}$  for some  $k \geq 1$ . If  $F$  is Poisson and  $\mathcal{R}$  is a family of linear functions then this model is referred to as Poisson (linear) INGARCH, see [Ferland et al. \(2006\)](#). Some authors also use the name Poisson autoregression for cases where  $F$  is Poisson and  $r$  could be nonlinear, see, e.g., [Fokianos and Tjøstheim \(2009\)](#), [Fokianos and Tjøstheim \(2012\)](#), [Fokianos \(2012\)](#), for the case  $p = q = 1$ . If  $q = 0$  then the model has a purely autoregressive structure and is abbreviated as INARCH( $p$ ). In the following the acronym PAR( $p$ ) is used for INARCH( $p$ ) with  $F$  Poisson and  $r$  linear. It has been shown that if  $p = q = 1$  and  $r$  is linear, i.e.,  $r(\mathbf{x}, \boldsymbol{\eta}) = \eta_1 + \eta_2 x_1 + \eta_3 x_2$ , and if  $\eta_1 \geq 0$  and  $\eta_2 + \eta_3 < 1$  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 (1), see [Davis and Liu \(2016\)](#). For an overview of conditions for strict stationarity and ergodicity for other choices of  $F$  see e.g., ([Ahmad and Francq 2016](#), Section 3).

A different class of models consists of integer autoregressive moving average (INARMA) models. These models arise from the same structure as the classical ARMA time series models, but the multiplication sign is replaced by the Steutel and van Harn’s thinning operator  $\circ$ . If  $Y$  is a count random variable and  $\alpha \in (0, 1)$  then

$$\alpha \circ Y := \sum_{i=1}^Y U_i, \tag{2}$$

where  $\{U_i\}$  are iid Bernoulli variables with  $\alpha = P(U_i = 1) = 1 - P(U_i = 0)$ , which are independent of  $Y$ , with the convention that an empty sum (the case  $Y = 0$ ) equals 0. Let  $\{\varepsilon_t\}$  be a sequence of iid count random variables with distribution  $G$  with a finite variance, and let  $\alpha_i, \beta_j \in (0, 1)$  for  $i = 1, \dots, p, j = 1, \dots, q$ . The INARMA( $p, q$ ) model is defined as

$$Y_t = \sum_{i=1}^p \alpha_i \circ Y_{t-i} + \varepsilon_t + \sum_{j=1}^q \beta_j \circ \varepsilon_{t-j}, \tag{3}$$

where the Bernoulli variables involved in all the thinning operations are independent and independent of  $\{\varepsilon_t\}$ . If  $p = 1, q = 0$  then the model (3) corresponds to the INAR(1) model introduced in McKenzie (1985) and Al-Osh and Alzaid (1987), taking the following form

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t = \sum_{i=1}^{Y_{t-1}} U_{t,i} + \varepsilon_t. \quad (4)$$

For  $\alpha \in (0, 1)$  there exists a strictly stationary solution of (4) and the law of the innovations  $\{\varepsilon_t\}$  uniquely determines the marginal distribution of  $Y_t$  as well as the conditional distribution  $Y_t | \mathcal{F}_{t-1}$ . In particular, if  $\{\varepsilon_t\}$  are iid Poisson then  $Y_t$  has Poisson distribution as well, and this special case has been considered in many applications. Model (3) with  $p > 1, q = 0$ , was introduced and studied by Du and Li (1991) and since then, many authors have considered variants of model (3), various extensions and modifications, see Scotto et al. (2015) for a comprehensive review.

Various GOF tests have been suggested in the literature for the aforementioned two classes of models. Neumann (2011) and Fokianos and Neumann (2013) considered GOF tests for the regression function  $r$  in a Poisson INGARCH(1,1); see model (1) with  $p = q = 1$  and Poisson  $F$ . A slightly less formal assessment of model adequacy is explored in Aleksandrov and Weiss (2020) for a PAR(1) model as well as for a Poisson INAR(1). GOF tests based on the sample index of dispersion were considered in Schweer and Weiss (2014) and Weiss et al. (2019) for a Poisson INAR(1), and by Weiss and Schweer (2015) for a PAR(1). A test based on the classical Pearson's  $\chi^2$  statistic for the marginal distribution specified by the null hypothesis is proposed by Weiss (2018b) for a Poisson INAR( $p$ ).

A different approach to GOF testing for time series of counts is based on the probability generating function (PGF). Recall that if  $Y$  is a nonnegative discrete random variable then its PGF is defined as  $g_Y(u) = E u^Y$  for all  $u$  for which the expectation is finite, which is always the case for  $u \in [0, 1]$ . The distribution of  $Y$  can be easily obtained from  $g_Y^{(k)}(0)$  and thus the PGF uniquely determines the distribution of  $Y$ . If  $\mathbf{Y} = (Y_1, Y_2)^\top$  is a count random vector then its PGF is defined as  $g_{\mathbf{Y}}(\mathbf{u}) = E(u_1^{Y_1} u_2^{Y_2})$  and exists for all  $\mathbf{u} = (u_1, u_2)^\top \in [0, 1]^2$ . Test criteria involving the estimates of the PGF of  $(Y_t, Y_{t-1})^\top$  have proved as useful not only for assessing the specification of  $F$  in (1) or  $G$  in (4), but also for determination the model itself, see Meintanis and Karlis (2014) for INAR(1) and Schweer (2016) for a more general setup involving INAR(1) and PAR(1) as special cases. In both mentioned articles the considered test statistic is an integrated  $L_2$  distance between the empirical PGF of  $(Y_t, Y_{t-1})^\top$  and the parametric estimate of  $g_{(Y_t, Y_{t-1})^\top}$  under the null hypothesis. Hudecová et al. (2015) also consider tests based on PGF, but instead of the estimator  $g_{(Y_t, Y_{t-1})^\top}$ , their criteria are constructed as an integrated  $L_2$  distance between the nonparametric estimate  $\hat{g}_n(u) = n^{-1} \sum_{t=1}^n u^{Y_t}$  of the marginal PGF  $g_Y$  and a semiparametric estimate of  $g_Y$ , which is model-specific. In the following sections we extend this approach from the univariate to the bivariate setup.

### 3. Bivariate Models for Time Series of Counts

The bivariate INAR and PAR models considered in this paper are based on the following bivariate distribution: We say that  $\mathbf{Y} = (Y_1, Y_2)^\top$  follows a bivariate Poisson distribution, see Kocherlakota and Kocherlakota (1992), denoted as  $BP(\lambda_1, \lambda_2, \phi)$ , if its PGF is

$$g_{\mathbf{Y}}(\mathbf{u}) = \exp\{\lambda_1(u_1 - 1) + \lambda_2(u_2 - 1) + \phi(u_1 - 1)(u_2 - 1)\}, \quad \mathbf{u} \in [0, 1]^2, \quad (5)$$

where  $\lambda_1 - \phi > 0, \lambda_2 - \phi > 0, \phi > 0$ . This distribution arises via the trivariate reduction method, i.e.,  $Y_1 = U_1 + U_2, Y_2 = U_1 + U_3$ , where  $U_1, U_2, U_3$  are independent with Poisson distribution with mean  $\phi, \lambda_1 - \phi, \lambda_2 - \phi$  respectively.

The models based on this kind of bivariate Poisson distribution seem to be the most popular in the literature despite some of their limitations. Please note that a different construction of a bivariate distribution with Poisson marginals is considered in [Lakshminarayana et al. \(1999\)](#).

### 3.1. Bivariate INAR Model

Suppose that we have a time series  $\{\mathbf{Y}_t\}$  composed by a pair of counts, i.e.,  $\mathbf{Y}_t := (Y_{t,1}, Y_{t,2})^\top$ . Following [Latour \(1997\)](#) we say that  $\{\mathbf{Y}_t\}$  follows a bivariate INAR model of the first order, in the following referred to as BINAR, if

$$\mathbf{Y}_t = \mathbf{A} \circ \mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{6}$$

where  $\boldsymbol{\varepsilon}_t = (\varepsilon_{t,1}, \varepsilon_{t,2})^\top$  are iid bivariate count random vectors with finite covariance matrix,  $\mathbf{A}$  denotes a  $2 \times 2$  matrix with elements  $a_{ij}, i, j = 1, 2$ , and the operator  $\circ$  is a multivariate generalization of (2). Namely the operator  $\circ$  from (6) acts on a count random vector  $\mathbf{Y}$  of dimension two by means of the equation

$$\mathbf{A} \circ \mathbf{Y} = \begin{pmatrix} a_{11} \circ Y_1 + a_{12} \circ Y_2 \\ a_{21} \circ Y_1 + a_{22} \circ Y_2 \end{pmatrix},$$

where the univariate thinning operators was defined in (2).

[Latour \(1997\)](#) showed that if the spectral radius (the absolute value of the largest eigenvalue) of  $\mathbf{A}$  is smaller than 1 and  $\boldsymbol{\varepsilon}_t$  has a finite covariance matrix then there exists a strictly stationary and ergodic process satisfying (6). Furthermore, the conditional least squares (CLS) estimate is shown to be consistent and asymptotically normal. Maximum likelihood estimation is considered in [Pedeli and Karlis \(2011\)](#) and [Pedeli and Karlis \(2013c\)](#), with a special emphasis on the Poisson specification for  $\boldsymbol{\varepsilon}_t$ . For estimation of BINAR with negative Binomial innovations the reader is referred to [Mamode Khan et al. \(2019\)](#) and references therein.

### 3.2. Bivariate PAR Model

Let  $\mathcal{F}_t = \sigma\{\mathbf{Y}_s, s \leq t\}$ . We say that  $\{\mathbf{Y}_t\}$  follows a bivariate Poisson autoregression model of the first order, referred to as BPAR, model if,

$$\mathbf{Y}_t | \mathcal{F}_{t-1} \sim \text{BP}(\lambda_{t,1}, \lambda_{t,2}, \phi), \quad \lambda_t = \begin{pmatrix} \lambda_{t,1} \\ \lambda_{t,2} \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}}_{\boldsymbol{\alpha}} + \underbrace{\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}}_{\mathbf{B}} \begin{pmatrix} Y_{t-1,1} \\ Y_{t-1,2} \end{pmatrix}, \tag{7}$$

where  $\alpha_1, \alpha_2 \geq 0$  and the matrix  $\mathbf{B}$  has non-negative entries and is of full-rank. This model is sometimes referred to as a bivariate Poisson INARCH(1) model, e.g., [Lee et al. \(2018\)](#). [Liu \(2012\)](#) proved that if  $2^{1-1/p} \|\mathbf{B}\|_p < 1$  for some  $1 \leq p \leq \infty$  then  $\{\mathbf{Y}_t, \lambda_t\}$  is strictly stationary and ergodic, with  $\|\mathbf{B}\|_p = \sup_{\mathbf{x} \neq 0, \mathbf{x} \in \mathbb{C}^2} \|\mathbf{B}\mathbf{x}\|_p / \|\mathbf{x}\|_p$ . Strong consistency of the conditional maximum likelihood estimator (CMLE) of parameters was proved by [Andreassen \(2013\)](#) under the extra assumption  $\phi < \min\{\alpha_1, \alpha_2\}$ , while [Lee et al. \(2018\)](#) further showed that the CMLE has an asymptotic normal distribution.

## 4. Goodness-of-Fit Tests

Assume that we have observations  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ , which come from a stationary bivariate time series of counts and we would like to perform a GOF test to a particular model for this data, with the model being fixed apart from finite-dimensional parameters. Let  $v(\cdot, \cdot)$  be a non-negative weight function which will be further specified below. We propose as test statistic the weighted distance measure

$$S_{n,v} = n \int_0^1 \int_0^1 [\hat{g}_n(u_1, u_2) - \hat{g}_{n,0}(u_1, u_2)]^2 v(u_1, u_2) du_1 du_2, \tag{8}$$

where  $\hat{g}_n(u_1, u_2)$  is a non-parametric estimate of  $g_Y(u_1, u_2)$  based on  $Y_1, \dots, Y_n$ , given by  $\hat{g}_n(u_1, u_2) = n^{-1} \sum_{t=1}^n u_1^{Y_{t,1}} u_2^{Y_{t,2}}$  and where  $\hat{g}_{n,0}(u_1, u_2)$  is a semi-parametric estimate of the PGF  $g_Y(u_1, u_2)$  of  $Y$ , which is constructed specifically under the model being tested. The null hypothesis is rejected for large values of  $S_{n,v}$ .

By analogy to most, if not all, of the previously published work we consider  $\mathbf{u} \in [0, 1]^2$  as our working interval, despite the fact that uniqueness of PGFs and corresponding consistency of the test might require working on a region of  $\mathbf{u}$  containing the origin. Nevertheless we further investigate this aspect of our tests by simulations. See Esquível (2008) for a recent account of uniqueness of PGFs.

Although the idea of the proposed test statistics is analogous as for the univariate models mentioned in Section 2, the extension to a bivariate case is not straightforward, neither in terms of asymptotics nor on computational grounds as the multivariate nature of the data brings in some elevated technical difficulties. We state the necessary assumptions and provide a formal proof for the asymptotics of the suggested test statistic. In addition, extension of our GOF tests to more complex models is briefly discussed. In fact, and although here we only treat the bivariate case, extension to higher dimension also comes completely natural with our methods.

#### 4.1. Tests for the BINAR Model

Let  $\mathcal{G} = \{g_\epsilon(\cdot; \theta) : \theta \in \Theta\}$  denote a parametric family of PGFs indexed by a parameter  $\theta \in \Theta$ , with  $\Theta$  being a compact subset of  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ . We would like to test the null hypothesis

$$\mathcal{H}_0 : \{Y_t\} \text{ follows model (6) for some sequence } \epsilon_t \text{ with PGF in } \mathcal{G}.$$

Assume that  $\{Y_t\}$  is strictly stationary with PGF  $g_Y$ . Then it follows from the properties of BINAR that

$$\begin{aligned} g_Y(u_1, u_2) &= E\left(u_1^{Y_{t,1}} u_2^{Y_{t,2}}\right) = E\left(E\left(u_1^{Y_{t,1}} u_2^{Y_{t,2}} \mid \mathcal{F}_{t-1}\right)\right) \\ &= g_\epsilon(u_1, u_2; \theta) \cdot g_Y(w_{11}w_{21}, w_{12}w_{22}), \end{aligned} \tag{9}$$

where

$$w_{ij} = w_{ij}(\mathbf{u}, \mathbf{a}) = 1 + a_{ij}(u_i - 1),$$

and  $\mathbf{a} := \text{vec}(\mathbf{A}) = (a_{11}, a_{21}, a_{12}, a_{22})^\top$  denotes the vectorized version of  $\mathbf{A}$ .

Here we used the fact that for a Binomial distribution with parameters  $(m, p)$  the PGF is  $(1 + p(u - 1))^m$ , so that if additionally  $\hat{\mathbf{a}}$  and  $\hat{\theta}$  are suitable estimators of the unknown parameters, then (9) yields a semi-parametric estimate of  $g_Y$  under model (6) as

$$\hat{g}_{n,0}(u_1, u_2) = g_\epsilon(\mathbf{u}, \hat{\theta}) \hat{g}_n(\hat{w}_{11}\hat{w}_{21}, \hat{w}_{12}\hat{w}_{22}), \tag{10}$$

where  $\hat{w}_{ij} = w_{ij}(\mathbf{u}, \hat{\mathbf{a}})$ .

We will consider a GOF test for the null hypothesis  $\mathcal{H}_0$  with  $\mathcal{G}$  being the family of bivariate Poisson distributions with unspecified parameters. Interest in testing  $\mathcal{H}_0$  lies in the fact that while this distribution is by far the most popular in the univariate as well as in the multivariate context, alternative specifications have also been employed such as models with innovations following a bivariate negative Binomial distribution; see for instance Pedeli and Karlis (2013a), Popović et al. (2018), and Kim and Lee (2017). Also BPAR models (see next section) may be considered to be alternative models of interest.

#### 4.2. Tests for the BPAR Model

The corresponding null hypothesis for the BPAR model is formulated as

$$\tilde{\mathcal{H}}_0 : \{Y_t\} \text{ follows model (7)}.$$

If  $\{\mathbf{Y}_t\}$  is strictly stationary, then it follows from the properties of the model that the PGF of  $\mathbf{Y}_t$  is given by

$$\begin{aligned}
 g_{\mathbf{Y}}(u_1, u_2) &= \mathbb{E}\left(u_1^{Y_{t,1}} u_2^{Y_{t,2}}\right) = \mathbb{E}\left(\mathbb{E}\left(u_1^{Y_{t,1}} u_2^{Y_{t,2}} \mid \mathcal{F}_{t-1}\right)\right) & (11) \\
 &= \mathbb{E}\left(\exp\{\lambda_{t,1}(u_1 - 1) + \lambda_{t,2}(u_2 - 1) + \phi(u_1 - 1)(u_2 - 1)\}\right) \\
 &= e^{\phi(u_1-1)(u_2-1) + \alpha_1(u_1-1) + \alpha_2(u_2-1)} \\
 &\quad \times \mathbb{E}\left(e^{Y_{t-1,1}[b_{11}(u_1-1) + b_{12}(u_2-1)]} e^{Y_{t-1,2}[b_{21}(u_1-1) + b_{22}(u_2-1)]}\right) \\
 &= e^{\phi(u_1-1)(u_2-1) + \alpha_1(u_1-1) + \alpha_2(u_2-1)} g_{\mathbf{Y}}(w_1, w_2), & (12)
 \end{aligned}$$

where

$$w_i = e^{b_{1i}(u_1-1) + b_{2i}(u_2-1)}, \quad i = 1, 2. \tag{13}$$

Thus, if we have an appropriate estimator  $\hat{\boldsymbol{\theta}}$  of the parameter  $\boldsymbol{\theta} := (\boldsymbol{\alpha}^\top, \text{vec}(\mathbf{B}), \phi)^\top = (\alpha_1, \alpha_2, b_{11}, b_{21}, b_{12}, b_{22}, \phi)^\top$ , then in view of (11) we may define a semi-parametric estimate of  $g_{\mathbf{Y}}$  as

$$\hat{g}_{n,0}(u_1, u_2) = e^{\hat{\phi}(u_1-1)(u_2-1) + \hat{\alpha}_1(u_1-1) + \hat{\alpha}_2(u_2-1)} \hat{g}_n(\hat{w}_1, \hat{w}_2) := c(\mathbf{u}, \hat{\boldsymbol{\theta}}) \hat{g}_n(\hat{w}_1, \hat{w}_2), \tag{14}$$

where

$$\hat{w}_i = e^{\hat{b}_{1i}(u_1-1) + \hat{b}_{2i}(u_2-1)},$$

and

$$c(\mathbf{u}, \boldsymbol{\theta}) = e^{\phi(u_1-1)(u_2-1) + \alpha_1(u_1-1) + \alpha_2(u_2-1)}. \tag{15}$$

Deviations from the null hypothesis  $\tilde{\mathcal{H}}_0$  include non-Poisson conditionals (see Heinen and Rengifo (2007)) as well as model violations towards more general specifications.

**Remark 1.** We should remark that the tests proposed in this article are not for BINAR or BPAR models per se. Specifically the test criterion for BINAR is for bivariate counts for which the PGF satisfies Equation (9), and analogously the test criterion for BPAR is for bivariate counts for which the PGF satisfies Equation (12). In the context of time series of counts however, BINAR and BPAR models are framed by equations (9) and (12), respectively, to such an extend that for all practical purposes these equations may be regarded as characterizing equations for the models themselves. Thus, our tests could be viewed as being on an equal footing with universally consistent methods such as those suggested by Fokianos and Neumann (2013), Jiménez-Gamero et al. (2020), and Leucht et al. (2015).

### 4.3. Computations

From (8) and by means of (10) and (14), we have after straightforward algebra

$$S_{n,v} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \Delta_{i,j},$$

where  $\Delta_{i,j} = \Delta(\mathbf{Y}_i, \mathbf{Y}_j)$  with

$$\Delta_{i,j} = \int_0^1 \int_0^1 D_i(u_1, u_2) D_j(u_1, u_2) v(u_1, u_2) du_1 du_2 \tag{16}$$

where

$$D_i(u_1, u_2) = u_1^{Y_{i,1}} u_2^{Y_{i,2}} - g_\varepsilon(u_1, u_2; \hat{\boldsymbol{\theta}}) (\hat{w}_{11} \hat{w}_{21})^{Y_{i,1}} (\hat{w}_{12} \hat{w}_{22})^{Y_{i,2}},$$

for the BINAR model, and

$$D_i(u_1, u_2) = u_1^{Y_{i,1}} u_2^{Y_{i,2}} - c(\mathbf{u}, \hat{\boldsymbol{\theta}}) \hat{w}_1^{Y_{i,1}} \hat{w}_2^{Y_{i,2}},$$

for the BPAR model.

The integral figuring in Equation (16) with weight function  $v(u_1, u_2) = u_1^\gamma u_2^\gamma$ ,  $\gamma \geq 0$ , may be expressed in more elementary terms. However the final fully reduced explicit forms, which are available from the authors upon request, are not convenient from the computational point of view. Thus, the tests were implemented by numerical evaluation of the corresponding integrals.

### 5. Asymptotics

In this section, we study the asymptotic distribution of the test statistic  $S_{n,v}$  under the null hypothesis of a BINAR model as well as the corresponding limit null distribution under a BPAR model. Results for the behavior under fixed alternatives are also provided and show that in both cases the test is consistent against certain fixed alternatives. Finally resampling versions are proposed that circumvent the problem of unknown components in the aforementioned limit null distributions. Please note that under the standing assumptions these results are valid for general innovation distribution and weight function as well as for arbitrary parameter-estimates.

#### 5.1. Asymptotics of the Test Statistic: BINAR Case

Assume the test statistic  $S_{n,v}$  (8) is constructed for testing  $\mathcal{H}_0$  with specified  $\mathcal{G} = \{g_\varepsilon(\cdot; \theta) : \theta \in \Theta\}$ , for a compact  $\Theta$ , i.e., the semi-parametric estimate  $\hat{g}_{n,0}$  is constructed by (10). Consider the following assumptions:

- (A.1) Let  $v(\cdot, \cdot)$  be a non-negative function satisfying  $0 < \int_0^1 \int_0^1 v(u_1, u_2) du_1 du_2 < \infty$ .
- (A.2) Let  $\{\mathbf{Y}_t\}$  follow model (6), where the spectral radius of  $\mathbf{A}$  is smaller than 1.
- (A.3) Let  $\mathcal{G}$  correspond to distributions with finite second moment. Furthermore, let the second partial derivatives of  $g_\varepsilon$  with respect to  $\theta$  exist and be continuous in  $\theta$  and suppose that

$$\begin{aligned} \sup_{\mathbf{u} \in [0,1]^2, \theta \in \Theta} \left| \frac{\partial g_\varepsilon(\mathbf{u}, \theta)}{\partial \theta_i} \right| &< \infty, \\ \sup_{\mathbf{u} \in [0,1]^2, \theta \in \Theta} \left| \frac{\partial}{\partial u_i} \frac{\partial g_\varepsilon(\mathbf{u}, \theta)}{\partial \theta_j} \right| &< \infty, \\ \int_{[0,1]^2} \left| \frac{\partial^2 g_\varepsilon(\mathbf{u}, \theta)}{\partial \theta_i \partial \theta_j} \right| v(\mathbf{u}) d\mathbf{u} &< \infty \end{aligned}$$

hold true for all  $\theta \in \Theta$  and  $i, j \in \{1, \dots, k\}$ , where (recall) that  $k$  is the dimension of  $\theta$ .

- (A.4) Let  $\hat{\zeta} := (\hat{\theta}^\top, \hat{\mathbf{a}}^\top)^\top$  be an estimator of  $\zeta := (\theta^\top, \mathbf{a}^\top)^\top$  such that for some  $q \geq 1$

$$\sqrt{n}(\hat{\zeta} - \zeta) = \frac{1}{\sqrt{n}} \sum_{t=q+1}^n \mathbf{1}(\mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots, \mathbf{Y}_{t-q}, \zeta) + o_P(1),$$

where  $\mathbf{1}_t = \mathbf{1}(\mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots, \mathbf{Y}_{t-q}, \zeta) = (\mathbf{1}_{t,1}^\top, \mathbf{1}_{t,2}^\top)^\top$  form a strictly stationary and ergodic sequence of martingale differences with finite variance. Here,  $\mathbf{1}_{t,1}$  is of the same dimension as  $\theta$  and  $\mathbf{1}_{t,2}$  has the dimension of  $\text{vec}(\mathbf{A})$ , i.e., four.

Under (A.2) and (A.3),  $\{\mathbf{Y}_t\}$  is strictly stationary and ergodic with finite second order moments, see Franke and Rao (1995). Regarding possible estimators  $\hat{\zeta}$  in (A.4), Franke and Rao (1995) considered the CMLE and proved its asymptotic normality under a set of regularity conditions. These conditions involve the finiteness of  $E(\|\varepsilon_t\|^3)$  and some further assumptions on the distribution of  $\varepsilon_t$ .

**Theorem 1.** Under (A.1)–(A.4) and as  $n \rightarrow \infty$ , the limit distribution of  $S_{n,v}$  under the null hypothesis  $\mathcal{H}_0$  is the same as the distribution of

$$\int_{[0,1]^2} Q^2(\mathbf{u}, \zeta) v(\mathbf{u}) d\mathbf{u},$$

where  $\{Q(\mathbf{u}, \zeta), \mathbf{u} \in [0, 1]^2\}$  is a Gaussian random field with zero mean and covariance structure

$$\begin{aligned} & E(Q(\mathbf{u}, \zeta)Q(\mathbf{v}, \zeta)) \\ &= E\left([u_1^{Y_{q+1,1}} u_2^{Y_{q+1,2}} - g_\varepsilon(\mathbf{u}, \boldsymbol{\theta})W_q(\mathbf{u}, \mathbf{a}) - \mathbf{l}_{q+1,1}(\zeta)^\top h_2(\mathbf{u}, \zeta) - \mathbf{l}_{q+1,2}(\zeta)^\top h_3(\mathbf{u}, \zeta)] \right. \\ & \quad \left. \times [v_1^{Y_{q+1,1}} v_2^{Y_{q+1,2}} - g_\varepsilon(\mathbf{v}, \boldsymbol{\theta})W_q(\mathbf{v}, \mathbf{a}) - \mathbf{l}_{q+1,1}(\zeta)^\top h_2(\mathbf{v}, \zeta) - \mathbf{l}_{q+1,2}(\zeta)^\top h_3(\mathbf{v}, \zeta)]\right), \end{aligned}$$

where  $\mathbf{l}_{q+1,1}$  and  $\mathbf{l}_{q+1,2}$  are from assumption (A.4) and

$$\begin{aligned} W_t(\mathbf{u}, \mathbf{a}) &= \prod_{i,j=1}^2 w_{ij}(\mathbf{u}, \mathbf{a})^{Y_{t,j}}, \\ h_2(\mathbf{u}, \zeta) &= \nabla_{\boldsymbol{\theta}} g_\varepsilon(\mathbf{u}, \boldsymbol{\theta}) g_Y(w_{11}w_{21}, w_{12}w_{22}), \\ h_3(\mathbf{u}, \zeta) &= g_\varepsilon(\mathbf{u}, \boldsymbol{\theta}) E(\nabla_{\mathbf{a}} W_1(\mathbf{u}, \mathbf{a})). \end{aligned}$$

The proof of the assertion is postponed to the Appendix A.

The asymptotic distribution of the test statistic  $S_{n,v}$  depends on several unknown quantities. One possibility is to generate the Gaussian random field figuring in Theorem 1 with the theoretical quantities replaced by some consistent estimators and then compute the critical values. Another possibility is the parametric bootstrap, which is quite natural here. The justification of the bootstrap approximation under the null hypothesis proceeds along similar lines as the proof of Theorem 1, conducted conditionally on the observed data  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  and with the help of assertions for triangular arrays and sums of martingale difference arrays.

Write  $S_n := S_n(\mathbf{Y}_1, \dots, \mathbf{Y}_n; \hat{\zeta})$  for the test statistic based on the original observations  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ , and the resulting parameter estimate  $\hat{\zeta}$ . (Here for simplicity we suppress the dependence of  $S_{n,v}$  on the weight function  $v(\cdot, \cdot)$ ).

1. Generate  $\varepsilon_t^*, t = 1, \dots, n$ , where  $\varepsilon_t^*$  are iid and follow the distribution with PGF  $g_\varepsilon(u, \hat{\boldsymbol{\theta}})$ .
2. Compute pseudo-observations  $\mathbf{Y}_{t'}^*, t = 1, \dots, n$ , using Equation (6) with  $\varepsilon_t$  replaced by  $\varepsilon_t^*, t = 1, \dots, n$ , and  $\mathbf{A}$  replaced by  $\hat{\mathbf{A}}$ .
3. Fit the model (6) using  $\mathbf{Y}_1^*, \dots, \mathbf{Y}_n^*$ , and compute the bootstrap estimator  $\hat{\zeta}^*$  of  $\zeta$ .
4. Compute the corresponding test statistic  $S_n^* := S_n(\mathbf{Y}_1^*, \dots, \mathbf{Y}_n^*; \hat{\zeta}^*)$ .
5. Repeat steps 1–4 several times, say  $B$ , and obtain the sequence of test statistics,  $S_{1,n}^*, \dots, S_{B,n}^*$ .
6. Compute the  $p$ -value as  $(\#b : S_{b,n}^* \geq S_n) / (B + 1)$ .

Next we shortly discuss the limit behavior of the test statistic under alternatives of type  $g_\varepsilon \notin \mathcal{G}$ . We assume that model (6) holds true but  $g_\varepsilon$  in the null hypothesis  $\mathcal{H}_0$  does not belong to  $\mathcal{G}$ . Moreover suppose that the estimators  $(\hat{\boldsymbol{\theta}}^\top, \hat{\mathbf{a}}^\top)^\top$  satisfy

$$(\hat{\boldsymbol{\theta}}^\top, \hat{\mathbf{a}}^\top)^\top \xrightarrow{P} (\boldsymbol{\theta}_A^\top, \mathbf{a}_A^\top)^\top, \tag{17}$$

for some  $\boldsymbol{\theta}_A \in \Theta$  and some  $\mathbf{a}_A$  such that the largest eigenvalue of the respective matrix  $A_A$  is in absolute value smaller than 1.



**Theorem 2.** Let  $\{\mathbf{Y}_t\}$  be strictly stationary with PGF  $g_A(u_1, u_2)$  that is continuous in  $(u_1, u_2) \in [0, 1]^2$  and let  $g_\varepsilon(\mathbf{u}, \boldsymbol{\theta})$  be continuous in  $\boldsymbol{\theta}$ . Let (A.1) and (17) be satisfied. Then, as  $n \rightarrow \infty$ ,

$$\frac{S_{n,v}}{n} \xrightarrow{P} \int_{[0,1]^2} \left\{ g_A(\mathbf{u}) - g_\varepsilon(\mathbf{u}, \boldsymbol{\theta}_A) g_A(w_{11,A} w_{21,A}, w_{12,A} w_{22,A}) \right\}^2 v(\mathbf{u}) d\mathbf{u}, \quad (18)$$

where  $w_{ij,A} = w_{ij}(\mathbf{u}, \mathbf{a}_A)$ ,  $i, j = 1, 2$ .

The proof is omitted since it suffices to follow the line of proof of Theorem 1 and use stationarity and ergodicity of  $\{\mathbf{Y}_t\}$ .

The right-hand side of (18) is strictly positive unless the true PGF  $g_\varepsilon(\cdot)$  coincides with the PGF  $g_\varepsilon(\cdot; \boldsymbol{\theta}_A)$  from the null hypothesis  $\mathcal{H}_0$ . This fact and the uniqueness of the PGF implies the consistency of the test which rejects the null hypothesis  $\mathcal{H}_0$  for large values of the test statistic  $S_{n,v}$  under such fixed alternatives. The test is also consistent for other types of fixed alternatives, e.g., against model violation. This feature of the test is further illustrated by Monte Carlo simulations in Section 7.

Please note that the test even has (non-negligible) power for some local alternatives, i.e., when the difference  $g_\varepsilon(\mathbf{u}) - g_\varepsilon(\mathbf{u}; \boldsymbol{\theta}_A)$  tends to 0 not too fast as  $n \rightarrow \infty$  and  $g_\varepsilon(\mathbf{u})$  depends on  $n$ . However, a rigorous proof of this result is quite technical, and therefore it is not discussed here any further.

### 5.2. Asymptotics of the Test Statistic: BPAR Case

This section considers the problem of testing the BPAR model. Recall that  $\boldsymbol{\theta} = (\alpha_1, \alpha_2, b_{11}, \dots, b_{22}, \phi)^\top$  stands for the vector of the parameters of the model (7), and suppose that:

- (B.1) The series  $\{\mathbf{Y}_t\}$  is strictly stationary solution of (7) with parameters  $\boldsymbol{\theta} \in \Theta$  such that  $\phi < \min\{\alpha_1, \alpha_2\}$  and  $\Theta$  is compact.
- (B.2) The estimator  $\hat{\boldsymbol{\theta}}$  of the parameter  $\boldsymbol{\theta}$  is such that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{t=q+1}^n \mathbf{l}(\mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots, \mathbf{Y}_{t-q}, \boldsymbol{\theta}) + o_P(1),$$

where  $\mathbf{l}_t = \mathbf{l}(\mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots, \mathbf{Y}_{t-q}, \boldsymbol{\theta}) = (l_{t,1}, \dots, l_{t,7})^\top$  form a strictly stationary and ergodic sequence of martingale differences with finite variances.

An estimator  $\hat{\boldsymbol{\theta}}$  that satisfies (B.2) is for instance the CMLE; see Lee et al. (2018).

**Theorem 3.** Under (A.1),(B.1)–(B.2) the limit distribution of  $S_{n,v}$  as  $n \rightarrow \infty$  is the same as the distribution of

$$\int_{[0,1]^2} Q^2(\mathbf{u}, \boldsymbol{\theta}) v(\mathbf{u}) d\mathbf{u},$$

where  $\{Q(\mathbf{u}, \boldsymbol{\theta}), \mathbf{u} \in [0, 1]^2\}$  is a Gaussian random field with zero mean and covariance structure

$$\begin{aligned} E(Q(\mathbf{u}, \boldsymbol{\theta})Q(\mathbf{v}, \boldsymbol{\theta})) = & E\left( [u_1^{Y_{t,1}} u_2^{Y_{t,2}} - c(\mathbf{u}, \boldsymbol{\theta}) w_1(\mathbf{u})^{Y_{t,1}} w_2(\mathbf{u})^{Y_{t,2}} + \mathbf{l}_t^\top \mathbf{d}(\boldsymbol{\theta}, \mathbf{u})] \right. \\ & \left. \times [v_1^{Y_{t,1}} v_2^{Y_{t,2}} - c(\mathbf{v}, \boldsymbol{\theta}) w_1(\mathbf{v})^{Y_{t,1}} w_2(\mathbf{v})^{Y_{t,2}} + \mathbf{l}_t^\top \mathbf{d}(\boldsymbol{\theta}, \mathbf{v})] \right), \end{aligned}$$

with  $\mathbf{l}_t$  defined in assumption (B.2),  $c(\mathbf{u}, \boldsymbol{\theta})$  defined in (15),

$$w_i(\mathbf{u}) = e^{\beta_{i1}(u_i-1) + \beta_{i2}(u_i-1)}, \quad i = 1, 2,$$

$$\mathbf{d}(\boldsymbol{\theta}, \mathbf{u}) = c(\mathbf{u}, \boldsymbol{\theta}) E\left( \mathbf{D}_1(\mathbf{u}) w_1^{Y_{1,1}} w_2^{Y_{1,2}} \right)$$

and

$$D_t(\mathbf{u}) = (u_1 - 1, u_2 - 1, (u_1 - 1)Y_{t,1}, (u_1 - 1)Y_{t,2}, (u_2 - 1)Y_{t,1}, (u_2 - 1)Y_{t,2}, (u_1 - 1)(u_2 - 1))^T.$$

Similarly as for the BINAR model, the parametric bootstrap can be carried out in a very natural way and is recommended for practical use. The justification of the bootstrap approximation would again proceed along similar lines as the proof of Theorem 3.

The following theorem is analogous to Theorem 2 and describes the behavior of the test statistic under some alternatives.

**Theorem 4.** Let  $\{Y_t\}$  be strictly stationary with PGF,  $\tilde{g}_A(u_1, u_2)$  that is continuous in  $(u_1, u_2) \in [0, 1]^2$  and let the estimator  $\hat{\theta}$  of  $\theta = (\alpha^\top, \text{vec}(\mathbf{B})^\top, \phi)^\top$  satisfy  $\hat{\theta} \xrightarrow{P} \theta_A$  for some  $\theta_A \in \Theta$ . Let (A.1) be satisfied. Then, as  $n \rightarrow \infty$ ,

$$\frac{S_{n,v}}{n} \xrightarrow{P} \int_{[0,1]^2} \left\{ \tilde{g}_A(\mathbf{u}) - c(\mathbf{u}, \theta_A) \tilde{g}_A(w_{1,A}, w_{2,A}) \right\}^2 v(\mathbf{u}) d\mathbf{u},$$

where  $w_{1,A}, w_{2,A}$  are defined by (13) with  $b_{ij}$  replaced by  $b_{ij,A}$ , and  $c(\mathbf{u}, \theta)$  is defined in (15).

The proof is omitted since it suffices to follow the line of the proof of Theorem 3 and to use stationarity and ergodicity of  $\{Y_t\}$ . Possible comments for this theorem are quite parallel to Theorem 1 and therefore are omitted.

### 6. Extension to Generalized BPAR Model

Extension of the proposed test to arbitrary higher order or dimension will be discussed in Section 9. In this section we aim at a new PGF-based GOF test for a relatively mild, yet very important, generalization of the BPAR model. Specifically consider the following model for  $\{Y_t\}$ :

$$Y_t | \mathcal{F}_{t-1} \sim \text{BP}(\lambda_{t,1}, \lambda_{t,2}, \phi), \quad \lambda_t = \alpha + \mathbf{A}\lambda_{t-1} + \mathbf{B}Y_{t-1}, \tag{19}$$

where  $\alpha = (\alpha_1, \alpha_2)^\top$  with  $\alpha_1, \alpha_2 \geq 0$ ,  $\mathbf{A}, \mathbf{B}$  are matrices of non-negative entries and  $\mathbf{B}$  is of full rank. (Please note that if  $\mathbf{A}$  is replaced by the zero matrix, then model (19) reduces to the BPAR model defined in (7)). Model (19) was studied, e.g., in Andreassen (2013), Liu (2012), and Lee et al. (2018), with the acronym INGARCH(1,1). If  $\|\mathbf{A}\|_p + 2^{1-1/p}\|\mathbf{B}\|_p < 1$  for some  $1 \leq p \leq \infty$ , then  $\{Y_t, \lambda_t\}$  is strictly stationary and ergodic, see Liu (2012), and if  $\phi < \min\{a_1, a_2\}$ , where  $(a_1, a_2)^\top = (\mathbf{I} - \mathbf{A})^{-1}\alpha$ , then the CMLE is consistent and asymptotically normal, see Andreassen (2013), Lee et al. (2018).

Assume that  $\{Y_t\}$  follows a stationary model (19). Following analogous arguments as in Section 4.2 we have that the PGF of  $\mathbf{Y}$  under this model is given by,

$$\begin{aligned} g_{\mathbf{Y}}(u_1, u_2) &= E\left(u_1^{Y_{t,1}} u_2^{Y_{t,2}}\right) = E\left(E\left(u_1^{Y_{t,1}} u_2^{Y_{t,2}} | \mathcal{F}_{t-1}\right)\right) \\ &= E\left(\exp\{\lambda_{t,1}(u_1 - 1) + \lambda_{t,2}(u_2 - 1) + \phi(u_1 - 1)(u_2 - 1)\}\right) \\ &= e^{\alpha_1(u_1-1) + \alpha_2(u_2-1) + \phi(u_1-1)(u_2-1)} E\left(w_{1,b}^{Y_{t-1,1}} w_{2,b}^{Y_{t-1,2}} w_{1,a}^{\lambda_{t-1,1}} w_{2,a}^{\lambda_{t-1,2}}\right) \\ &= e^{\alpha_1(u_1-1) + \alpha_2(u_2-1) + \phi(u_1-1)(u_2-1)} h_{\mathbf{Y},\lambda}(w_{1,b}, w_{2,b}, w_{1,a}, w_{2,a}), \end{aligned} \tag{20}$$

where

$$w_{i,a} = e^{a_{1i}(u_1-1) + a_{2i}(u_2-1)}, \quad w_{i,b} = e^{b_{1i}(u_1-1) + b_{2i}(u_2-1)}, \quad i = 1, 2, \tag{21}$$

with the joint PGF of the vector  $(\mathbf{Y}^\top, \lambda^\top)^\top$  defined by

$$h_{\mathbf{Y},\lambda}(u_1, u_2, v_1, v_2) := E\left(u_1^{Y_{t,1}} u_2^{Y_{t,2}} v_1^{\lambda_{t,1}} v_2^{\lambda_{t,2}}\right).$$

It is straightforward to devise a test for the model (19) by using Equation (20) and proceeding analogously as with Equation (8). However the asymptotics of such a test as well as its actual implementation require a separate investigation. In this connection preliminary Monte Carlo results showed some promise but there were also problems, and therefore we decide not to pursue this extension any further here.

## 7. Simulations

The finite sample behavior of the suggested bootstrap test is investigated in the following simulation study. We consider the null hypotheses  $\mathcal{H}_0$  of Poisson BINAR and  $\widetilde{\mathcal{H}}_0$  of BPAR model and investigate the size of the test under the null hypothesis and the power for various alternatives.

The unknown parameters of the BINAR model are estimated using the CLS method, and  $\phi$  is estimated using the moment method, see Pedeli and Karlis (2013b). The parameters of BPAR model are estimated by the CMLE method. For simplicity the weight function is set to  $v(\cdot, \cdot) \equiv 1$ , i.e., we take  $\gamma = 0$  in  $v(u_1, u_2) = u_1^\gamma u_2^\gamma$ . The simulations were conducted in the R-computing environment R Core Team (2019) and by employing the warp-speed bootstrap of Giacomini et al. (2013) for  $M = 1000$  repetitions. When using this method,  $B=1$  bootstrap samples are generated for each Monte Carlo repetition and the resulting p-value is computed from the overall bootstrap sample of  $M$  replicas.

Our results are for sample size  $n = 100, 200$  and  $n = 500$  at level of significance  $\alpha = 0.01, 0.05$  and  $0.1$  for both  $\mathcal{H}_0$  and  $\widetilde{\mathcal{H}}_0$ . For the BINAR model, a reasonable alternative might be BPAR model or model (6) with innovations  $\varepsilon_t$  following a distribution other than the bivariate Poisson. Such a popular alternative is a bivariate distribution with negative Binomial marginals. There are several possibilities as to how to generate such variables. Here we consider the bivariate negative Binomial distribution  $\text{BNB}(\lambda_1, \lambda_2, r)$  of Dunn (1967), whereby  $(U_1, U_2)^\top \sim \text{BNB}(\lambda_1, \lambda_2, r)$ , with  $U_i, i = 1, 2$ , being marginally negative Binomial with mean  $\lambda_i$  and variance  $\lambda_i(1 + \lambda_i/r)$  and  $\text{cov}(U_1, U_2) = \lambda_1\lambda_2/r$ . This bivariate negative Binomial distribution is also used in an alternative considered to a BPAR model. Namely a model of form (7) with conditional distribution  $\text{BNB}(\lambda_{t1}, \lambda_{t2}, r)$  instead of bivariate Poisson is considered, with the dependence of  $\lambda_{ti}$  on  $\mathcal{F}_{t-1}$  the same as specified in Equation (7). We will refer to this model as the negative Binomial BINARCH. Finally, we also explored the power of the test when testing  $\widetilde{\mathcal{H}}_0$  and the data follow BINAR model and vice versa.

The results for the size of the two tests are summarized in Table 1 and for the BINAR model were obtained by using either  $\mathbf{A} = \mathbf{A}_1$  or  $\mathbf{A} = \mathbf{A}_2$ , where

$$\mathbf{A}_1 = \begin{pmatrix} 0.3 & 0.1 \\ 0.2 & 0.2 \end{pmatrix}, \mathbf{A}_2 = \begin{pmatrix} 0.6 & 0.0 \\ 0.0 & 0.7 \end{pmatrix},$$

and  $\varepsilon_t$  following a  $\text{BP}(6, 4, 3)$  distribution. For BPAR, we set  $\alpha = (6, 4)^\top$ ,  $\phi = 3$ , and again  $\mathbf{B}$  equal either to  $\mathbf{A}_1$  or  $\mathbf{A}_2$ . As it may be seen from Table 1, the tests are conservative if the matrix  $\mathbf{A}_1$  is used in the data generating process. On the other hand, the size slightly exceeds the nominal size  $\alpha$  if  $\mathbf{A}_2$  is used in the data generating process. Please note that a similar phenomenon was already observed for the univariate series in Hudecová et al. (2015): The test may be slightly conservative for certain parametric settings of the model at hand, while it can be slightly anticonservative for other settings. Recall also that the unknown parameters are estimated by a CLS method for BINAR models whereas CMLE is used for BPAR models, which might also partly explain the different behaviour observed in Table 1. In either case however, the observed size approaches the prescribed significance level as the sample size increases.

In this connection, the rather poor small sample behavior observed in the left part of Table 1 for the BINAR model parameterised by the matrix  $\mathbf{A}_1$ , can be improved by considering a modified test statistic for which integration in Equation (8) is carried over  $[-1, 1]^2$  rather than over  $[0, 1]^2$ ; see the corresponding discussion at the end of Section 4.

Specifically this approach (performed on the same simulated data) and for nominal size  $\alpha = 0.05$  leads to empirical size equal to 0.033, 0.039 and 0.052, for  $n = 100, 200$ , and 500, respectively. For all other settings however, the results for the modified test statistic (available from the authors upon a request) are very similar and hence we only present here results for the original test statistic  $S_{n,v}$  defined by Equation (8).

**Table 1.** Size of the test for testing Poisson BINAR in  $\mathcal{H}_0$  and BPAR in  $\widetilde{\mathcal{H}}_0$ .

$n \backslash \alpha$	Poisson BINAR						BPAR					
	A = A <sub>1</sub>			A = A <sub>2</sub>			B = A <sub>1</sub>			B = A <sub>2</sub>		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
100	0.000	0.010	0.030	0.011	0.039	0.104	0.002	0.029	0.075	0.013	0.077	0.145
200	0.001	0.010	0.048	0.017	0.065	0.114	0.006	0.040	0.076	0.028	0.094	0.168
500	0.002	0.025	0.069	0.005	0.049	0.105	0.005	0.050	0.094	0.014	0.053	0.127

The power of the test for  $\mathcal{H}_0$  with data from a BINAR model with negative Binomial innovations is provided in Table 2. The considered bivariate negative Binomial distribution is BNB(6, 4,  $r$ ) for  $r = 5, 10$ . Please note that for larger  $r$ , the BNB distribution is closer to a bivariate Poisson distribution, and this fact is also reflected in the power of the test, which is lower for  $r = 10$ . For example, a sample size  $n = 100$  seems to be insufficient for distinguishing between Poisson and negative binomial INAR for this larger  $r$ . However, as the sample size  $n$  grows, the test performs very well even for  $r = 10$  irrespective of the matrix **A** being used to simulate the model.

**Table 2.** Power of the test for  $\mathcal{H}_0$  for data from BINAR model with negative Binomial marginals.

$n \backslash \alpha$	A = A <sub>1</sub>						A = A <sub>2</sub>					
	r = 5			r = 10			r = 5			r = 10		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
100	0.005	0.139	0.328	0.004	0.065	0.135	0.357	0.581	0.694	0.080	0.196	0.296
200	0.120	0.639	0.902	0.009	0.119	0.331	0.715	0.897	0.955	0.167	0.432	0.540
500	0.951	1.000	1.000	0.132	0.884	0.984	1.000	1.000	1.000	0.637	0.861	0.920

A similar observation also holds for the results in Table 3 which correspond to the null hypothesis  $\widetilde{\mathcal{H}}_0$  with data from a negative Binomial BINARCH. These results were obtained with the same values of  $\alpha$  and **B** used for the null hypothesis, and  $r = 30, 50$ .

**Table 3.** Power of the test for  $\widetilde{\mathcal{H}}_0$  with data generated by a negative Binomial BINARCH model.

$n \backslash \alpha$	B = A <sub>1</sub>						B = A <sub>2</sub>					
	r = 30			r = 50			r = 30			r = 50		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
100	0.085	0.357	0.469	0.060	0.169	0.265	0.347	0.672	0.821	0.124	0.350	0.526
200	0.446	0.740	0.833	0.143	0.345	0.460	0.448	0.871	0.956	0.203	0.568	0.746
500	0.942	0.989	0.994	0.502	0.701	0.797	0.810	0.996	1.000	0.524	0.874	0.957

On the other hand, Table 4 reports the power of the test for  $\mathcal{H}_0$  with data following a BPAR model and the power of the test for  $\widetilde{\mathcal{H}}_0$  with data from a Poisson BINAR model. The results in Table 4 show that if the matrix **B** in the BPAR model equals **A**<sub>1</sub> then the test for  $\mathcal{H}_0$  fails to distinguish between the two models for sample sizes up to  $n = 500$ . In contrast, if the matrix **B** is equal to **A**<sub>2</sub> in BPAR and we test for Poisson BINAR then the power is satisfactory even for  $n = 200$ . The same observation holds for the opposite situation when one tests for BPAR and the data come from a Poisson BINAR.

**Table 4.** Power of the test of Poisson BINAR in  $\mathcal{H}_0$  for data following BPAR and vice versa.

$n \backslash \alpha$	Test $\mathcal{H}_0$ for Data from BPAR						Test $\widetilde{\mathcal{H}}_0$ for Data from BINAR(1)					
	$\mathbf{B} = \mathbf{A}_1$			$\mathbf{B} = \mathbf{A}_2$			$\mathbf{A} = \mathbf{A}_1$			$\mathbf{A} = \mathbf{A}_2$		
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
100	0.003	0.022	0.061	0.613	0.791	0.852	0.003	0.024	0.082	0.011	0.244	0.626
200	0.006	0.027	0.102	0.955	0.988	0.994	0.004	0.058	0.168	0.109	0.810	0.942
500	0.009	0.136	0.278	1.000	1.000	1.000	0.030	0.151	0.346	0.724	0.980	0.995

## 8. Real-Data Application

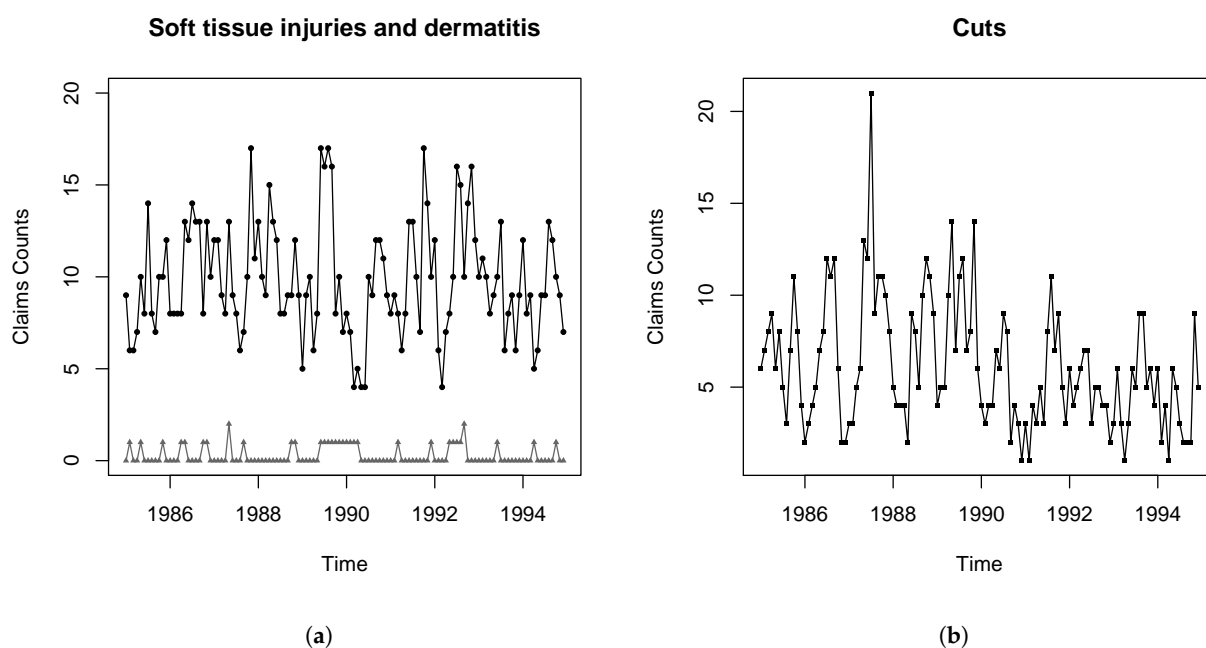
Joint modelling of count observations finds important applications in the insurance industry, see for instance Partrat (1994), and Vernic (1997). In this connection, it is a common practice for insurance companies to split the reported claims into several types. Typically, it is reasonable to expect that aggregate amounts (daily or monthly) of these different types of claims to be dependent, see, e.g., Shi et al. (2016). If the mean size of claims counts is high, then classical models for continuous variables could be applied. On the other hand, if the observed counts of claims are formed by small integers it is appropriate to treat the data as genuine counts, and, consequently, engage models and methods specifically tailored for count time series.

We illustrate this kind of application on real data sets on the monthly number of claims of short-term disability benefits made by injured workers to the British Columbia Workers Compensation Board. The time period is from January 1985 to December 1994. The original data set from Freeland (1998) contains five time series corresponding to five different injury categories: burn injuries, soft tissue injuries, cuts, dermatitis and dislocations. These five time series have been previously analyzed by several authors, and separate univariate models were fitted. It has been found that the Poisson INAR is appropriate for all five series, except for series #3 (cuts), for which this model is not appropriate, see e.g., Freeland (1998), Zhu and Joe (2006), Hudecová et al. (2015). In particular, Freeland and McCabe (2004) and Zhu and Joe (2006) suggest to model the time series of cuts claims using an extension of INAR(1) model with a seasonal component. On the other hand, Biswas and Song (2009) argue that the ACF and PACF *do not indicate a significant seasonal effect*.

We first consider a bivariate series of soft tissue injuries claims and dermatitis claims; see Figure 1 (left panel). Possible dependence among these two series may be due to the fact that a major accident causes several injuries, often of different types. Previous analyses in Freeland (1998) or Hudecová et al. (2015) reveal that the marginal INAR(1) models might be appropriate for these two series. Hence, we consider a Poisson BINAR model with a diagonal matrix  $\mathbf{A}$  and estimate the parameters using the conditional least squares method (note that if a general, i.e., non-diagonal, matrix  $\mathbf{A}$  is considered, then the estimates of the off-diagonal entries are very close to zero). Estimators of parameters were obtained by the methods used in the simulations of the previous section. The matrix  $\mathbf{A}$  is estimated as  $\hat{\mathbf{A}} = \text{diag}(0.452, 0.293)$ , and the parameters of the bivariate Poisson distribution of the innovations  $\varepsilon_t$  are estimated as  $\hat{\lambda} = (5.376, 0.202)^\top$ ,  $\hat{\phi} = 0.109$ . The resulting GOF test applied with  $v \equiv 1$  and  $B = 999$  bootstrap samples leads to  $S_{n,v} = 8.618 \cdot 10^{-4}$  with  $p$ -value 0.327. Thus, one may conclude that the Poisson BINAR model with a diagonal matrix  $\mathbf{A}$  seems to fit the data well.

On the other hand, if one considers a bivariate series of soft tissue injuries claims and cuts claims, see Figure 1 (right panel), and tests the GOF for a Poisson BINAR model with a diagonal matrix  $\mathbf{A}$  (which corresponds to univariate INAR models for the two series), the null hypothesis is rejected with  $p$ -value 0.0370. A Poisson BINAR with a general matrix  $\mathbf{A}$  is rejected as well ( $p$ -value 0.002). This is in accordance with the findings of previous papers mentioned above. However, the hypothesis of BPAR is not rejected with  $p$ -value 0.277. The MLEs in this model are  $\hat{\alpha} = (5.144, 2.944)^\top$ ,  $\hat{\phi} = 0.721$  and  $\hat{\mathbf{B}}$  is such that  $\text{vec}(\hat{\mathbf{B}}) = (0.455, 0.004, 0.043, 0.521)^\top$ . One could further scrutinize this data set by postulating an extended (nonstationary) BPAR model with a seasonal component. This

might improve the models' predictions compared to the fitted stationary model, but such a model lies outside the scope of this paper.



**Figure 1.** Monthly number of claims. (a) Soft tissue injuries claims (black dots) and dermatitis claims (gray triangles). (b) Cuts claims.

## 9. Concluding Remarks

We suggest consistent goodness-of-fit tests for bivariate INAR and bivariate Poisson autoregressive models, estimated by least squares and maximum likelihood, respectively, with well defined limit null distributions. Since these limit distributions are complicated we suggest parametric bootstrap resampling which engages distributional assumptions featuring in the null hypothesis in order to actually carry out the tests. Monte Carlo results show that this bootstrap version of the new tests is generally reasonably sized and has good power against certain popular alternative configurations. Our real-data applications are in the direction of further scrutiny and better understanding of the mechanisms generating the data at hand.

At this point, we wish to discuss potential extension of the tests to models of higher order or dimension, such as the multivariate INAR type processes studied by [Franke and Rao \(1995\)](#), [Latour \(1997\)](#), [Pedeli and Karlis \(2011\)](#), [Pedeli and Karlis \(2013c\)](#), and [Pedeli and Karlis \(2013a\)](#), and corresponding extensions of PAR type models considered by [Liu \(2012\)](#), [Andreassen \(2013\)](#), [Lee et al. \(2018\)](#), [Ciu and Zhu \(2018\)](#) and [Ciu et al. \(2020\)](#). In this connection, and while it is conceptually straightforward to extend the tests for INAR or PAR models of higher order, see also Section 6 of [Hudecová et al. \(2015\)](#), we wish to emphasize the fact that a certain order and/or dimension increase brings about serious challenges in estimation as well as on the actual implementation of the methods due to the potentially great number of new parameters introduced.

Finally, before we close, we wish to point out that while this paper deals solely with stationary time series models, certain real world phenomena, including time series of counts, often exhibit deterministic trends or seasonal patterns, and thus GOF tests for non-stationary models would be of great practical interest. However, despite the fact that univariate INAR and PAR models with deterministic components have been extensively studied in the literature, multivariate extensions of such nonstationary models are rather scarce; we refer to [Santos et al. \(2019\)](#) for some recent works.

**Author Contributions:** Conceptualization, Š.H., M.H., S.G.M.; methodology, Š.H., M.H., S.G.M.; software Š.H.; writing—original draft preparation, Š.H., S.G.M.; writing—review and editing, Š.H., M.H., S.G.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work of Šárka Hudecová and Marie Hušková is funded by the Czech Science Foundation project GA18-08888S.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available in [Freeland \(1998\)](#) or on request from the corresponding author.

**Acknowledgments:** The authors are grateful to the Associated Editor and the anonymous reviewers for their careful inspection of the paper and valuable comments, which led to an improved manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Proofs

#### Appendix A.1. Proof of Theorem 1.

**Proof.** Denote as  $\mathcal{A} \subset [0, 1]^4$  such that if  $\mathbf{a} \in \mathcal{A}$  then  $\mathbf{A}$  has the largest eigenvalue smaller than 1. The test statistic  $S_{n,v}$  is of the form

$$\begin{aligned} S_{n,v} &= \int_{[0,1]^2} \left[ \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( u_1^{Y_{t,1}} u_2^{Y_{t,2}} - g_\varepsilon(\mathbf{u}, \hat{\boldsymbol{\theta}}) W_t(\mathbf{u}, \hat{\mathbf{a}}) \right) \right]^2 v(\mathbf{u}) d\mathbf{u} \\ &= \int_{[0,1]^2} \left[ \frac{1}{\sqrt{n}} \sum_{t=2}^n \left( u_1^{Y_{t,1}} u_2^{Y_{t,2}} - g_\varepsilon(\mathbf{u}, \hat{\boldsymbol{\theta}}) W_{t-1}(\mathbf{u}, \hat{\mathbf{a}}) \right) \right]^2 v(\mathbf{u}) d\mathbf{u} + o_P(1). \end{aligned}$$

Let  $A_t = A_t(\boldsymbol{\theta}, \mathbf{a}) = g_\varepsilon(\mathbf{u}, \boldsymbol{\theta}) W_t(\mathbf{u}, \mathbf{a})$ . In the following, we omit for a moment the argument  $\mathbf{u}$  in all the considered functions. Namely we consider  $W_t = W_t(\mathbf{a})$  as a function of the argument  $\mathbf{a}$  alone, i.e.,  $\nabla W_t$  is a differential of  $W_t$  with respect to  $\mathbf{a}$ ,  $\nabla W_t = \nabla_{\mathbf{a}} W_t = (\partial W / \partial a_{11}, \partial W / \partial a_{12}, \partial W / \partial a_{21}, \partial W / \partial a_{22})^\top$ . Likewise,  $\nabla g_\varepsilon = \nabla_{\boldsymbol{\theta}} g_\varepsilon$  is a differential of  $g_\varepsilon$  with respect to  $\boldsymbol{\theta}$ . Then a Taylor expansion gives:

$$A_t(\hat{\boldsymbol{\theta}}, \hat{\mathbf{a}}) = A_t(\boldsymbol{\theta}, \mathbf{a}) + \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \hat{\mathbf{a}} - \mathbf{a} \end{pmatrix}^\top \begin{pmatrix} W_t(\mathbf{a}) \nabla g_\varepsilon(\boldsymbol{\theta}) \\ g_\varepsilon(\boldsymbol{\theta}) \nabla W_t(\mathbf{a}) \end{pmatrix} + R_t,$$

where

$$R_t = \frac{1}{2} \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \hat{\mathbf{a}} - \mathbf{a} \end{pmatrix}^\top \begin{pmatrix} W_t(\mathbf{a}^*) H_{g_\varepsilon}(\boldsymbol{\theta}^*) & \nabla g_\varepsilon(\boldsymbol{\theta}^*) \nabla W_t(\mathbf{a}^*)^\top \\ \nabla W_t(\mathbf{a}^*) \nabla g_\varepsilon(\boldsymbol{\theta}^*)^\top & g_\varepsilon(\boldsymbol{\theta}^*) H_{W_t}(\mathbf{a}^*) \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \hat{\mathbf{a}} - \mathbf{a} \end{pmatrix},$$

with  $\mathbf{a}^* = \hat{\mathbf{a}} + c(\mathbf{a} - \hat{\mathbf{a}})$  and  $\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}} + d(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$  for some  $c, d \in (0, 1)$  and where  $H_{g_\varepsilon} = (\partial^2 g_\varepsilon / (\partial \theta_k \partial \theta_l))_{k,l}$  is the  $k \times k$  matrix of second partial derivatives of  $g_\varepsilon$  with respect to  $\boldsymbol{\theta}$ . Similarly,  $H_{W_t}$  is the  $4 \times 4$  matrix of second partial derivatives of  $W_t$  with respect to  $\mathbf{a}$ . For  $\nabla W_t$  we have

$$\frac{\partial W_t}{\partial a_{ij}} = \frac{W_t}{w_{ij}^{Y_{t,j}}} Y_{t-1,j} w_{ij}^{Y_{t,j}-1} (u_i - 1) = W_t \frac{Y_{t,j}}{w_{ij}} (u_i - 1),$$

and clearly  $|\partial W_t / \partial a_{ij}| \leq Y_{tj}$  for all  $\mathbf{u} \in [0, 1]^2$  and  $\mathbf{a} \in \mathcal{A}$ . Furthermore,

$$\frac{\partial^2 W_t}{\partial a_{ij} \partial a_{kl}} = \begin{cases} W_t \frac{Y_{t-1,j} Y_{t-1,l}}{w_{ij} w_{kl}} (u_i - 1)(u_k - 1), & (i, j) \neq (k, l), \\ W_t \frac{Y_{t-1,j} (Y_{t-1,j} - 1)}{w_{ij}^2} (u_i - 1)^2 & (i, j) = (k, l), \end{cases}$$

and thus,  $|\partial^2 W_t / (\partial a_{ij} \partial a_{kl})| \leq Y_{t-1,j} Y_{t-1,l}$  for all  $\mathbf{u} \in [0, 1]^2$  and all  $\mathbf{a} \in \mathcal{A}$ .

Due to the finite second order moments of  $\{\mathbf{Y}_t\}$  and Assumption (A.2) we have

$$\left| \frac{1}{\sqrt{n}} \sum_{t=2}^n R_t \right| \leq K_1 \sqrt{n} \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 + \sqrt{n} \|\hat{\mathbf{a}} - \mathbf{a}\|^2 K_2 + K_3 \sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathbf{1}_{k \times 4} (\hat{\mathbf{a}} - \mathbf{a}),$$

where  $K_i$  are constants and  $\mathbf{1}_{k \times 4}$  is a matrix of 1's of dimension  $k \times 4$ . Hence, it follows from the Cauchy–Schwartz inequality that  $\int_{[0,1]^2} \left| \frac{1}{\sqrt{n}} \sum_{t=2}^n R_t \right|^2 v(\mathbf{u}) d\mathbf{u} \xrightarrow{P} 0$ , as  $n \rightarrow \infty$ , and thus the asymptotic distribution of  $S_{n,v}$  is the same as the asymptotic distribution of

$$\int_{[0,1]^2} [J_{1n}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) + J_{2n}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) + J_{3n}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u})]^2 v(\mathbf{u}) d\mathbf{u},$$

where

$$\begin{aligned} J_{1n}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) &= \frac{1}{\sqrt{n}} \left( \sum_{t=2}^n u_1^{Y_{t,1}} u_2^{Y_{t,2}} - g_\varepsilon(\mathbf{u}, \boldsymbol{\varepsilon}) W_{t-1}(\mathbf{u}, \mathbf{a}) \right), \\ J_{2n}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) &= -\sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \nabla g_\varepsilon(\mathbf{u}, \boldsymbol{\theta}) \frac{1}{n} \sum_{t=2}^n W_{t-1}(\mathbf{u}, \mathbf{a}), \\ J_{3n}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) &= -g_\varepsilon(\mathbf{u}, \boldsymbol{\theta}) \sqrt{n} (\hat{\mathbf{a}} - \mathbf{a})^\top \frac{1}{n} \sum_{t=2}^n \nabla_{\mathbf{a}} W_{t-1}(\mathbf{u}, \mathbf{a}). \end{aligned}$$

Regarding the behavior of  $J_{2n}$ , notice that  $|W_t(\mathbf{a})| \leq 1$  and thus, for a given  $\mathbf{a}$  and as  $n \rightarrow \infty$

$$\frac{1}{n} \sum_{t=2}^n W_{t-1}(\mathbf{a}) \xrightarrow{P} E W_1(\mathbf{u}, \mathbf{a}) = E \prod_{i,j=1}^2 w_{ij}^{Y_{1,j}} = g_Y(w_{11} w_{21}, w_{12} w_{22}),$$

uniformly in  $\mathbf{u} \in [0, 1]^2$  due to the uniform ergodicity theorem. Now define

$$J_{2n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) = - \left[ \frac{1}{\sqrt{n}} \sum_{t=q+1}^n \mathbf{1}_{t1} \right]^\top h_2(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}),$$

where  $h_2(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) = \nabla_{\boldsymbol{\theta}} g_\varepsilon(\boldsymbol{\theta}, \mathbf{u}) g_Y(w_{11} w_{21}, w_{12} w_{22})$ . Due to (A.2), it follows that  $\int_{[0,1]^2} [J_{2n}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) - J_{2n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u})]^2 v(\mathbf{u}) d\mathbf{u} = o_P(1)$  as  $n \rightarrow \infty$ . Similarly, in view of the form of  $\nabla W_{t-1}(\mathbf{a})$  given above and the finiteness of  $E Y_{1j}$ , we get from the uniform ergodicity theorem that as  $n \rightarrow \infty$

$$g_\varepsilon(\boldsymbol{\theta}, \mathbf{a}) \frac{1}{n} \sum_{t=2}^n \nabla W_{t-1}(\mathbf{u}, \mathbf{a}) \xrightarrow{P} g_\varepsilon(\mathbf{u}, \boldsymbol{\theta}) E \nabla W_1(\mathbf{u}, \mathbf{a}) =: h_3(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u})$$

uniformly in  $\mathbf{u} \in [0, 1]^2$ . Define

$$J_{3n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) = -g_\varepsilon(\boldsymbol{\theta}, \mathbf{u}) \left[ \frac{1}{\sqrt{n}} \sum_{t=q+1}^n \mathbf{1}_{t2} \right]^\top h_3(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}).$$



As  $|h_3(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u})| \leq \max_{j=1,2} EY_{1,j}$  then  $\int_{[0,1]^2} [J_{3n}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) - J_{3n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u})]^2 v(\mathbf{u}) d\mathbf{u} = o_P(1)$ . Hence, it suffices to study the asymptotic behavior of  $\int_{[0,1]^2} Q_n^2(\mathbf{u}, \boldsymbol{\theta}, \mathbf{a}) v(\mathbf{u}) d\mathbf{u}$ , where

$$Q_n(\mathbf{u}, \boldsymbol{\theta}, \mathbf{a}) = J_{1n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) + J_{2n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) + J_{3n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u})$$

with  $J_{1n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) = \frac{1}{\sqrt{n}} \left( \sum_{t=q+1}^n u_1^{Y_{t,1}} u_2^{Y_{t,2}} - g_\epsilon(\mathbf{u}, \boldsymbol{\theta}) W_{t-1}(\mathbf{u}, \mathbf{a}) \right)$ . We will make use Theorem 22 from Ibragimov and Chasminskij (1981) and the subsequent remark in order to show that the integral  $\int_{[0,1]^2} Q_n^2(\mathbf{u}, \boldsymbol{\theta}, \mathbf{a}) v(\mathbf{u}) d\mathbf{u}$  converges in distribution to  $\int_{[0,1]^2} Q^2(\mathbf{u}, \boldsymbol{\theta}, \mathbf{a}) v(\mathbf{u}) d\mathbf{u}$ , where  $Q$  is a Gaussian random field specified in the statement of the theorem. To this end, we need to verify that

I.  $\sup_{n, \mathbf{u}, \boldsymbol{\theta}, \mathbf{a}} E Q_n^2(\mathbf{u}, \boldsymbol{\theta}, \mathbf{a}) < \infty$ .

II. There exist constants  $\alpha > 0$  and  $H > 0$  such that

$$\sup_{n, \boldsymbol{\theta}, \mathbf{a}} E |Q_n(\mathbf{u}, \boldsymbol{\theta}, \mathbf{a}) - Q_n(\mathbf{v}, \boldsymbol{\theta}, \mathbf{a})|^2 \leq H \|\mathbf{u} - \mathbf{v}\|^\alpha.$$

III. The marginal distributions of  $Q_n$  converge to the marginal distributions of  $Q$  uniformly in  $\boldsymbol{\theta}$  and  $\mathbf{a}$ .

First, notice that  $Q_n$  is of a form  $\frac{1}{\sqrt{n}} \sum_{t=q+1}^n L_t$ , where  $\{L_t\}$  is a martingale difference sequence  $L_t = u_1^{Y_{t,1}} u_2^{Y_{t,2}} - g_\epsilon(\boldsymbol{\theta}, \mathbf{u}) W_{t-1}(\mathbf{u}, \mathbf{a}) - \mathbf{I}_{t1}(\boldsymbol{\theta}, \mathbf{a})^\top h_2(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) - \mathbf{I}_{t2}(\boldsymbol{\theta}, \mathbf{a})^\top h_3(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u})$ . Thus  $E Q_n^2 \leq \frac{1}{n} \sum_{t=q+1}^n E |L_t|^2 \leq 4(1 + g_\epsilon^2 + \|h_2\|^2 E \|\mathbf{I}_{11}\|^2 + \|h_3\|^2 E \|\mathbf{I}_{12}\|^2)$ . Since we assume finite variances of  $\mathbf{I}_{t,j}$  and thus  $\|h_j\|^2$  are bounded (due to finiteness of  $EY_{t,j}$  and assumption A.2), condition I. directly follows. Condition III. follows from the central limit theorem for martingale difference sequences on  $L_t$ .

In order to prove condition II., we will show that  $E |J_{jn0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) - J_{jn0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{v})|^2 \leq D \|\mathbf{u} - \mathbf{v}\|^\alpha$  for  $j = 1, 2, 3$ , for some  $D > 0$  and  $\alpha > 0$ . In the following  $K$  is a generic constant. For  $j = 1$  we have

$$\begin{aligned} & E |J_{1n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) - J_{1n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{v})|^2 \\ &= \frac{1}{n} E \left\{ \sum_{t=q+1}^n u_1^{Y_{t,1}} u_2^{Y_{t,2}} - v_1^{Y_{t,1}} v_2^{Y_{t,2}} - g_\epsilon(\mathbf{u}, \boldsymbol{\theta}) W_{t-1}(\mathbf{u}, \mathbf{a}) + g_\epsilon(\mathbf{v}, \boldsymbol{\theta}) W_{t-1}(\mathbf{v}, \mathbf{a}) \right\}^2 \\ &\leq K \left\{ E [u_1^{Y_{t,1}} u_2^{Y_{t,2}} - v_1^{Y_{t,1}} v_2^{Y_{t,2}}]^2 + g_\epsilon(\mathbf{u}, \boldsymbol{\theta})^2 E [W_1(\mathbf{u}, \mathbf{a}) - W_1(\mathbf{v}, \mathbf{a})]^2 \right. \\ &\quad \left. + (g_\epsilon(\mathbf{u}, \boldsymbol{\theta}) - g_\epsilon(\mathbf{v}, \boldsymbol{\theta}))^2 E |W_1(\mathbf{u}, \mathbf{a})|^2 \right\}. \end{aligned}$$

By the mean value theorem applied on the function  $f(x, y) = x^a y^b$  and due to the fact that  $E |Y_{t,j}|$  is finite, we get that  $E [u_1^{Y_{t,1}} u_2^{Y_{t,2}} - v_1^{Y_{t,1}} v_2^{Y_{t,2}}]^2 \leq K \|\mathbf{u} - \mathbf{v}\|^2$ . Next,  $|g_\epsilon| \leq 1$  and the partial derivatives  $|\partial W_1 / \partial u_j|$  are bounded by  $|Y_{1,1} + Y_{1,2}|$ , which has finite expectation. This implies that  $g_\epsilon(\mathbf{u}, \boldsymbol{\theta})^2 E [W_1(\mathbf{u}, \mathbf{a}) - W_1(\mathbf{v}, \mathbf{a})]^2 \leq K \|\mathbf{u} - \mathbf{v}\|^2$ . Finally, it follows from the definition of  $g_\epsilon$  as a PGF that it is Lipschitz. Furthermore,  $E |W_1(\mathbf{u}, \mathbf{a})|^2 \leq 1$ , which implies  $(g_\epsilon(\mathbf{u}, \boldsymbol{\theta}) - g_\epsilon(\mathbf{v}, \boldsymbol{\theta}))^2 E |W_1(\mathbf{u}, \mathbf{a})|^2 \leq K \|\mathbf{u} - \mathbf{v}\|^2$ . Together, we get  $E |J_{1n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) - J_{1n0}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{v})|^2 \leq K \|\mathbf{u} - \mathbf{v}\|^2$ . Similar arguments are used to show that the condition holds also for  $j = 3$ . If we use the assumption that  $l_t$  has finite variances, it remains to show that  $\|h_j(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) - h_j(\boldsymbol{\theta}, \mathbf{a}, \mathbf{v})\|^2 \leq K \|\mathbf{u} - \mathbf{v}\|^\alpha$ . First, notice that the partial derivatives of  $h_3$  with respect to  $\mathbf{u}$  are continuous functions for  $\mathbf{u} \in [0, 1]^2$ . Thus, they are bounded and we can apply the mean value theorem on the components of  $h_3$ , which implies that  $\|h_3(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) - h_3(\boldsymbol{\theta}, \mathbf{a}, \mathbf{v})\|^2 \leq K \|\mathbf{u} - \mathbf{v}\|^2$ . Similarly, if the assumption (A.2) holds then the partial derivatives of  $h_2$  are bounded and this implies that  $\|h_2(\boldsymbol{\theta}, \mathbf{a}, \mathbf{u}) - h_2(\boldsymbol{\theta}, \mathbf{a}, \mathbf{v})\|^2 \leq K \|\mathbf{u} - \mathbf{v}\|^2$ , which completes the proof of condition II., and thus the assertion of the theorem follows.  $\square$

### Appendix A.2. Proof of Theorem 3.

**Proof.** The proof proceeds along the same lines as the proof of Theorem 1. A Taylor expansion is used to show that

$$S_{n,v} = \int_0^1 \int_0^1 [J_0(\boldsymbol{\theta}, \mathbf{u}) - J_1(\boldsymbol{\theta}, \mathbf{u})]^2 v(\mathbf{u}) d\mathbf{u} + o_P(1),$$

where  $\mathbf{D}_t(\mathbf{u})$  are defined in the statement of the theorem,

$$J_0(\boldsymbol{\theta}, \mathbf{u}) = \frac{1}{\sqrt{n}} \sum_{t=2}^n [u_1^{Y_{t,1}} u_2^{Y_{t,2}} - f_{t-1}(\boldsymbol{\theta}, \mathbf{u})],$$

$$J_1(\boldsymbol{\theta}, \mathbf{u}) = n^{-1/2} \sum_{t=2}^n f_{t-1}(\boldsymbol{\theta}, \mathbf{u}) \mathbf{D}_{t-1}(\mathbf{u})^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

$$f_{t-1}(\boldsymbol{\theta}, \mathbf{u}) = c(\mathbf{u}, \boldsymbol{\theta}) w_1(\mathbf{u})^{Y_{t-1,1}} w_2(\mathbf{u})^{Y_{t-1,2}}.$$

The term  $J_1(\boldsymbol{\theta}, \mathbf{u})$  can be replaced by  $J_{10}(\boldsymbol{\theta}, \mathbf{u}) = [\mathbf{E} \mathbf{D}_1(\mathbf{u})^\top f_1(\boldsymbol{\theta}, \mathbf{u})] (n^{-1/2} \sum_{t=q}^n \mathbf{1}_t)$ . Finally, the conditions I.–III. from Theorem 22 of [Ibragimov and Chasminskij \(1981\)](#) can be verified similarly as it is done for BINAR using the fact that the fourth moments of  $\mathbf{Y}_t$  are finite and the parametric space  $\Theta$  is compact.  $\square$

## References

- Ahmad, Ali, and Christian Francq. 2016. Poisson QMLE of count time series models. *Journal of Time Series Analysis* 37: 291–314. [\[CrossRef\]](#)
- Al-Osh, Mohamed A., and Abdulhamid A. Alzaid. 1987. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* 8: 261–75. [\[CrossRef\]](#)
- Aleksandrov, Boris, and Christian H. Weiss. 2020. Testing the dispersion structure of count time series using Pearson residuals. *ASTA Advances in Statistical Analysis* 104: 325–61. [\[CrossRef\]](#)
- Andreassen, Camilla Mondrup. 2013. Models and Inference for Correlated Count Data. Ph.D. thesis, Aarhus University, Aarhus, Denmark.
- Biswas, Atanu, and Peter Xue Kun Song. 2009. Discrete-valued ARMA processes. *Statistics and Probability Letters* 79: 1884–89. [\[CrossRef\]](#)
- Ciu, Yan, Qi Li, and Fukang Zhu. 2020. Flexible bivariate Poisson integer-valued GARCH model. *Annals of the Institute of Statistical Mathematics* 72: 1449–77.
- Ciu, Yan, and Fukang Zhu. 2018. A new bivariate integer-valued GARCH model allowing for negative cross-correlation. *TEST* 27: 428–52.
- Davis, Richard A., Scott H. Holan, Robert Lund, and Nalini Ravishanker. 2015. *Handbook of Discrete-Valued Time Series*. New York: Chapman and Hall/CRC.
- Davis, Richard A., and Heng Liu. 2016. Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica* 26: 1673–707. [\[CrossRef\]](#)
- Du, Jin-Guan, and Yuan Li. 1991. The integer-valued autoregressive (INAR( $p$ )) model. *Journal of Time Series Analysis* 12: 129–42.
- Dunn, James E. 1967. Characterization of the bivariate negative binomial distribution. *Journal of the Arkansas Academy of Science* 21: 77–86.
- Esquível, Manuel L. 2008. Probability generating functions for discrete real-valued random variables. *Theory of Probability and Its Applications* 52: 40–57. [\[CrossRef\]](#)
- Ferland, René, Alain Latour, and Driss Oraichi. 2006. Integer-valued GARCH processes. *Journal of Time Series Analysis* 27: 923–42. [\[CrossRef\]](#)
- Fokianos, Konstantinos. 2012. Count time series. In *Handbook of Statistics 30: Time Series—Methods and Applications*. Edited by Tata Subba Rao, Suhasini Subba Rao and Calyampudi Radhakrishna Rao. Amsterdam: Elsevier, pp. 315–47.
- Fokianos, Konstantinos, and Michael H. Neumann. 2013. A goodness-of-fit tests for Poisson count processes. *Electronic Journal of Statistics* 7:793–819. [\[CrossRef\]](#)
- Fokianos, Konstantinos, Andres Rahbek, and Dag Tjøstheim. 2009. Poisson autoregression. *Journal of the American Statistical Association* 104: 1430–39. [\[CrossRef\]](#)
- Fokianos, Konstantinos, and Dag Tjøstheim. 2012. Nonlinear Poisson autoregression. *Annals of the Institute of Statistical Mathematics* 64: 1205–25. [\[CrossRef\]](#)
- Franke, Jürgen, and T. Subba Rao. 1995. *Multivariate First-Order Integer-Valued Autoregressions*. Technical Report. Kaiserslautern: Universität Kaiserslautern.

- Freeland, R. Keith. 1998. Statistical Analysis of Discrete Time series with Application to the Analysis of Workers' Compensation Claims Data. Ph.D. thesis, Management Science Division, Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, BC, Canada. Available online: <https://open.library.ubc.ca/cIRcle/collections/ubctheses> (accessed on 25 February 2021).
- Freeland, R. Keith, and Brendan P. M. McCabe. 2004. Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis* 25: 701–22. [[CrossRef](#)]
- Giacomini, Raffaella, Dimitris Politis, and Halbert White. 2013. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* 29: 567–89. [[CrossRef](#)]
- Heinen, Andréas, and Erick Rengifo. 2007. Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance* 14: 564–583. [[CrossRef](#)]
- Hudecová, Šárka, Marie Hušková, and Simos G. Meintanis. 2015. Tests for time series of counts based on the probability generating function. *Statistics* 49: 316–37. [[CrossRef](#)]
- Ibragimov, Il'dar Abdulovich, and Rafail Zalmonovich Chasminskij. 1981. *Statistical Estimation, Asymptotic Theory*. New York: Springer.
- Jiménez-Gamero, M. Dolores, Sangyeol Lee, and Simos G. Meintanis. 2020. Goodness-of-fit tests for parametric specifications of conditionally heteroscedastic models. *TEST* 29: 682–703. [[CrossRef](#)]
- Kim, Hanwool, and Sangyeol Lee. 2017. On first order integer-valued autoregressive process with Katz family innovations. *Journal of Statistical Computation and Simulation* 87: 546–62. [[CrossRef](#)]
- Kocherlakota, Subrahmaniam, and Kathleen Kocherlakota. 1992. *Bivariate Discrete Distributions*. New York: Marcel Dekker Inc.
- Lakshminarayana, J., S. N. Narahari Pandit, and K. Srinivasa Rao. 1999. On a bivariate Poisson distribution. *Communications in Statistics—Theory and Methods* 28: 267–76. [[CrossRef](#)]
- Latour, Alain. 1997. The multivariate GINAR( $p$ ) process. *Advances in Applied Probability* 29: 228–48. [[CrossRef](#)]
- Lee, Youngmi, Sangyeol Lee, and Dag Tjøstheim. 2018. Asymptotic normality and parameter change test for bivariate Poisson INGARCH models. *TEST* 27: 52–69. [[CrossRef](#)]
- Leucht, Anne, Jens-Peter Kreiss, and Michael H. Neumann. 2015. A model specification test for GARCH(1,1) processes. *Scandinavian Journal of Statistics* 42: 1167–93. [[CrossRef](#)]
- Liu, Heng. 2012. Some Models for Time Series of Counts. Ph.D. thesis, Columbia University, New York, NY, USA.
- Mamode Khan, Naushad Ali, Yuvraj Sunecher, Vandna Jowaheer, Miroslav M. Ristić, and Maleika Heenaye–Mamode Khan. 2019. Investigating GQL-based inferential approaches for non-stationary BINAR(1) model under different quantum of over-dispersion with application. *Computational Statistics* 34: 1275–313. [[CrossRef](#)]
- McKenzie, Eddie. 1985. Some simple models for discrete variate time series. *Water Resources Bulletin* 21: 645–50. [[CrossRef](#)]
- McKenzie, Eddie. 2003. Discrete variate time series. *Stochastic Processes: Modelling and Simulation*. In *Handbook of Statistics*. Amsterdam: Elsevier, vol. 21, pp. 573–606.
- Meintanis, Simos G., and Dimitris Karlis. 2014. Validation tests for the innovation distribution in INAR time series models. *Computational Statistics* 29: 1221–41. [[CrossRef](#)]
- Neumann, Michael H. 2011. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* 17: 1258–84. [[CrossRef](#)]
- Partrat, Christian. 1994. Compound model for two dependent kinds of claim. *Insurance: Mathematics and Economics* 15: 219–31. [[CrossRef](#)]
- Pedeli, Xanthi, and Dimitris Karlis. 2011. A bivariate INAR(1) process with application. *Statistical Modelling* 11: 325–49. [[CrossRef](#)]
- Pedeli, Xanthi, and Dimitris Karlis. 2013a. On composite likelihood estimation of a multivariate INAR(1) model. *Journal of Time Series Analysis* 34: 206–20. [[CrossRef](#)]
- Pedeli, Xanthi, and Dimitris Karlis. 2013b. On estimation of the bivariate Poisson INAR process. *Communications in Statistics—Simulation and Computation* 42: 514–33. [[CrossRef](#)]
- Pedeli, Xanthi, and Dimitris Karlis. 2013c. Some properties of multivariate INAR(1) processes. *Computational Statistics & Data Analysis* 67: 213–25.
- Popović, Predrag M., Aleksandar S. Nastić, and Miroslav M. Ristić. 2018. Residual analysis with bivariate INAR models. *REVSTAT* 16: 349–64.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Santos, Cláudia, Isabel Pereira, and Manuel G. Scotto. 2019. On the theory of periodic multivariate INAR processes. *Statistical Papers* doi:10.1007/s00362-019-01136-5. [[CrossRef](#)]
- Schweer, Sebastian. 2016. A goodness-of-fit test for integer valued autoregressive processes. *Journal of Time Series Analysis* 37: 77–98. [[CrossRef](#)]
- Schweer, Sebastian, and Christian H. Weiss. 2014. Compound Poisson INAR(1) processes: Stochastic properties and testing for overdispersion. *Computational Statistics & Data Analysis* 77: 267–84.
- Scotto, Manuel G., Christian H. Weiss, and Sónia Gouveia. 2015. Thinning-based models in the analysis of integer-valued time series: A review. *Statistical Modelling* 15: 590–618. [[CrossRef](#)]
- Shi, Peng, Xiaoping Feng, and Jean-Philippe Boucher. 2016. Multilevel modeling of insurance claims using copulas. *The Annals of Applied Statistics* 10: 834–36. [[CrossRef](#)]
- Steutel, Fred W., and Klaas van Harn. 1979. Discrete analogues of self-decomposability and stability. *Annals of Probability* 7: 893–99. [[CrossRef](#)]

- 
- Vernic, Raluca. 1997. On the bivariate generalized poisson distribution. *ASTIN Bulletin* 27: 22–32. [[CrossRef](#)]
- Weiss, Christian H. 2018a. *Discrete-Valued Time Series*. Hoboken: John Wiley & Sons.
- Weiss, Christian H. 2018b. Goodness-of-fit testing of a count series' marginal distribution. *Metrika* 81: 619–51. [[CrossRef](#)]
- Weiss, Christian H., Annika Homburg, and Pedro Puig. 2019. Testing for zero inflation and overdispersion in INAR(1) models. *Statistical Papers* 60: 473–98. [[CrossRef](#)]
- Weiss, Christian H., and Sebastian Schweer. 2015. Detecting overdispersion in INARCH(1) processes. *Statistica Neerlandica* 69: 281–97. [[CrossRef](#)]
- Zhu, Rong, and Harry Joe. 2006. Modelling count data time series with Markov processes based on binomial thinning. *Journal of Time Series Analysis* 27: 725–38. [[CrossRef](#)]