TEST (2024) 33:1014–1040
https://doi.org/10.1007/s11749-024-00933-x

ORIGINAL PAPER

Specifications tests for count time series models with covariates

Šárka Hudecová[1] · Marie Hušková[1] · Simos G. Meintanis[2,3]

Received: 26 October 2023 / Accepted: 7 May 2024 / Published online: 1 July 2024

Abstract
We propose a goodness-of-fit test for a class of count time series models with covariates which includes the Poisson autoregressive model with covariates (PARX) as a special case. The test criteria are derived from a specific characterization for the conditional probability generating function, and the test statistic is formulated as a L_2 weighting norm of the corresponding sample counterpart. The asymptotic properties of the proposed test statistic are provided under the null hypothesis as well as under specific alternatives. A bootstrap version of the test is explored in a Monte–Carlo study and illustrated on a real data set on road safety.

Keywords Count time series with covariates · PARX · INGARCH-X · Goodness-of-fit test · Probability generating function · Bootstrap test

Mathematics Subject Classification 62M10 · 62F03

1 Introduction

Let $\{Y_t\}$ be a time series of counts, i.e., a time series of variables taking values in $\mathbb{N}_0$, and let $\{\mathbf{X}_t\}$ be a m-dimensional vector time series of exogenous covariates. Denote as I_t the information set available up to time t and let $\{\mathbf{Z}_t\}$ be a time series of d-

Š. Hudecová, M. Hušková, and S. G. Meintanis have contributed equally to this work.

✉ Simos G. Meintanis
simosmei@econ.uoa.gr

Šárka Hudecová
hudecova@karlin.mff.cuni.cz

Marie Hušková
huskova@karlin.mff.cuni.cz

1 Department of Probability and Mathematical Statistics, Charles University, Prague 186 75, Czech Republic

2 Department of Economics, National and Kapodistrian University of Athens, Athens, Greece

3 Pure and Applied Analytics, North–West University, Potchefstroom, South Africa

dimensional random vectors such that $\boldsymbol{Z}_t$ may contain past values of both endogenous and exogenous covariates. Our aim is to develop procedures for parametric specifications of the conditional distribution of Y_t given $\mathcal{I}_{t-1}$, i.e., we are interested in testing the null hypothesis

$$H_0 : Y_t | I_{t-1} \sim F_{\lambda_t}, \quad \lambda_t = h(\boldsymbol{Z}_t; \boldsymbol{\theta}), \tag{1}$$

where F_λ denotes a fixed count distribution with a mean λ, h is a specified function, and $\boldsymbol{\theta}$ is an unspecified vector of parameters. Notice that the manner in which λ_t depends on the covariate vector $\boldsymbol{Z}_t$ is fixed under H_0 apart from a finite-dimensional parameter $\boldsymbol{\theta}$ which must be estimated from the data.

We will restrict ourselves to a class of cumulative distribution functions $\{\mathcal{F}_\lambda\}$ such that $\lambda > 0$ is the mean of F_λ and the family $\{\mathcal{F}_\lambda\}$ satisfies

$$\lambda_1 \leq \lambda_2 \Rightarrow F_{\lambda_1}(y) \geq F_{\lambda_2}(y) \quad \text{for all } y \in \mathbb{R}. \tag{2}$$

Classes of distributions satisfying (2) are discussed in Aknouche and Francq (2021), where it is shown that the condition (2) holds, for instance, for Poisson distribution with mean λ, for Negative Binomial distribution with mean λ and fixed dispersion parameter ϕ, or for a zero-inflated Poisson distribution. Aknouche and Francq (2021) also provide conditions for existence of a stationary and ergodic solution to (1) with a conditional distribution satisfying (2).

If F_λ denotes the Poisson distribution with mean λ and if no exogenous variables are considered then (1) corresponds to the INGARCH model of (Fokianos and Tjøstheim 2009). An extension to this model, abbreviated sometimes as PARX, incorporating also exogenous variables was proposed and studied by Agosto et al. (2016), and the model has already found a number of important extensions and applications in various fields. In finance, Agosto et al. (2016) used PARX to model US corporate default counts, while Agosto and Raffinetti (2019) analyzed Italian corporate default counts data in the real estate sector. Other applications are in sports (Angelini and De Angelis 2017) and disease modelling (Agosto and Giudici 2020). Alternative specifications for $\mathcal{F}_\lambda$ include the Negative Binomial (NB) distribution (Christou and Fokianos 2014; Zhu 2011; Liboschik et al. 2017), and the zero-inflated Poisson distribution (Aknouche and Francq 2021), among others.

If the main objective of a data analysis is to consistently estimate the model parameters $\boldsymbol{\theta}$ (and so the conditional mean), then the conditional distribution $\mathcal{F}_\lambda$ does not have to be specified, because one can use quasi-maximum likelihood methods (QML) for the estimation, see Aknouche and Francq (2021) for the Poisson QML in this context. However, in many applications a probability forecast is of interest in a sense that the whole conditional distribution of future Y_{T+1} needs to be estimated given data up to time T. Then the correct specification of $\mathcal{F}_\lambda$ is crucial. It is, therefore, important to have a goodness-of-fit test that is sensitive to an incorrect specification of the conditional mean λ_t as well as to an incorrect conditional distribution $\mathcal{F}_\lambda$ in (1).

In problems when the whole underlying distribution is being tested, goodness-of-fit tests for continuous data typically employ the cumulative distribution function. When dealing with count distributions however, the use of a cumulative distribution function is often inconvenient due to its non-continuous nature; see for instance Gürtler and Henze (2000) where even in the case of testing for the Poisson distribution, the application of the Cramér–von Mises test is non-trivial. On the other hand, Gürtler and Henze (2000) propose the use of tests based on the probability generating function (PGF) as straightforward to apply and outperforming other more classical procedures. In this connection we wish to point out that PGF-based inference procedures, dating back to Bishop (1979), have proved quite efficient in a number of instances with i.i.d. count data. The reader is referred to Baringhaus and Henze (1992), and to Nakamura and Pérez-Abreu (1993) for review of earlier uses of the PGF. More recent references include 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. The theoretical underpinnings are that the PGF (see (Johnson et al. 2005, Chapter 1) and (Zayed 1996, Chapter 34)), uniquely determines the corresponding count distribution, while its empirical counterpart, the empirical PGF, is continuous in its argument, unbiased as estimator, and converges almost surely to the population PGF pointwise as well as uniformly on [0,1]; see for instance Rémillard and Theodorescu (2000), where weak convergence of the empirical PGF process on [0,1] is also shown.

The problem of testing goodness-of-fit within INGARCH models without covariates has been considered by many authors, for instance (Fokianos and Neumann 2013; Weiss and Schweer 2015; Schweer 2016; Hudecová et al. 2015, 2021) to mention some of them. Several of these procedures make use of the PGF and share much of the same favorable features as in the i.i.d. case. Unfortunately, generalization of these methods to models with exogenous covariates is not straightforward. We propose a method for testing the more general dynamic model specified in $\mathcal{H}_0$ for an arbitrary family of conditional distributions satisfying (2), a possibly nonlinear function h and a setup with external covariates. To the best of our knowledge, a goodness-of-fit test for this situation has not yet been investigated in the literature. Moreover and while our test is based on a characterization of the conditional mean by Bierens (1982), which has been successfully used in construction of various specification tests in different contexts, e.g., see (Escanciano and Velasco 2006; Hlávka et al. 2017) and references therein, and despite the fact of earlier uses of PGFs with related problems, our test formulation is new in that it combines Bierens' characterization and PGFs in a novel way not to be found in the existing literature. In this connection we also wish to refer to the recent article by Jiménez-Gamero et al. (2020) where this characterization is employed in order to test for correct specification of covariance in GARCH models, but this is a rather direct application of Bierens' result, whereas in the present work the incorporation of the population PGF within Bierens' characterization brought forward new computational and methodological challenges.

Hence, the novelty of theoretical and methodological contribution of this paper may be summarized as follows:

- We propose for a first time a goodness-of-fit test for an INGARCH model with external covariates. The test statistic is formulated by combining Bierens' characterization and the population (null) PGF in a novel way.
- The test is applicable to nonlinear models as well as non-Poisson models.

- The proposed test is sensitive to both types of model misspecification: An incorrectly specified conditional mean λ_t as well as incorrect conditional distribution $\mathcal{F}_{\lambda_t}$.

The rest of the paper is outlined as follows. Section 2 introduces the considered model. The test statistic is proposed in Sect. 3, while its asymptotic properties are studied in Sect. 4. Section 5 describes the practical computation of the test and presents a bootstrap version. An application to the Poisson autoregression with covariates (PARX) is discussed in Sect. 5.3. Section 6 presents results of a Monte Carlo study. A real data application of road safety is included in Sect. 7. We finally conclude with discussion of our finding and further outlook in Sect. 8. Proofs of all assertions are provided in an online Supplementary file.

2 Model specification

Let $p \geq 1$ and $q \geq 0$ be given integers. Let $\{\mathcal{F}_\lambda, \lambda \in \Omega \subset (0, \infty)\}$ be some family of discrete distributions on $\mathbb{N}_0$ indexed by a mean parameter $\lambda \in (0, \infty)$ and satisfying (2). Let $\boldsymbol{\Theta}$ be a subset of $\mathbb{R}^K$ for some $K \geq 1$ such that $\boldsymbol{\Theta} = \boldsymbol{\Theta}^{(1)} \times \boldsymbol{\Theta}^{(2)}$, where $\boldsymbol{\Theta}^{(i)} \subset \mathbb{R}^{K_i}$ for $K_i \geq 0$, $K_1 + K_2 = K$. Consider a model

$$Y_t | I_{t-1} \sim \mathcal{F}_{\lambda_t}, \quad \lambda_t = h(\boldsymbol{Z}_t; \boldsymbol{\theta}), \tag{3}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a vector of unknown model parameters such that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}^{(i)}$, $i = 1, 2$,

$$\boldsymbol{Z}_t = (Y_{t-1}, \ldots, Y_{t-p}, \lambda_{t-1}, \ldots, \lambda_{t-q}, \boldsymbol{X}_{t-1}^\top)^\top \tag{4}$$

is a vector of dimension $d = p+q+m$, which contains past values of both endogenous and exogenous covariates, and $h : [0, \infty)^{p+q} \times \mathbb{R}^{m+K} \to (0, \infty)$ is a link function such that

$$h(z; \boldsymbol{\theta}) = h(\boldsymbol{y}, \boldsymbol{l}, \boldsymbol{x}; \boldsymbol{\theta}) = r(\boldsymbol{y}, \boldsymbol{l}; \boldsymbol{\theta}_1) + \pi(\boldsymbol{x}; \boldsymbol{\theta}_2) \tag{5}$$

for $\boldsymbol{z} \in [0, \infty)^{p+q} \times \mathbb{R}^m$ such that $\boldsymbol{z} = (\boldsymbol{y}^\top, \boldsymbol{l}^\top, \boldsymbol{x}^\top)^\top$, $\boldsymbol{y} \in [0, \infty)^p$, $\boldsymbol{l} \in [0, \infty)^q$, $\boldsymbol{x} \in \mathbb{R}^m$ and for some non-negative functions $r : [0, \infty)^{p+q} \times \mathbb{R}^{K_1} \to (0, \infty)$ and $\pi : \mathbb{R}^{m+K_2} \to (0, \infty)$. This means that the conditional mean of Y_t given $\mathcal{I}_{t-1}$ is specified as

$$\mathbb{E}[Y_t | \mathcal{I}_{t-1}] = \lambda_t = r(Y_{t-1}, \ldots, Y_{t-p}, \lambda_{t-1}, \ldots, \lambda_{t-q}; \boldsymbol{\theta}_1) + \pi(\boldsymbol{X}_{t-1}; \boldsymbol{\theta}_2).$$

Notice that the random vector $\boldsymbol{Z}_t$ implicitly depends on the value of the model parameter $\boldsymbol{\theta}$, but we do not stress this dependence if not necessary. If $q = 0$ then the vector $\boldsymbol{Z}_t$ contains only past values of Y_t and $\boldsymbol{X}_t$, and the conditional mean is specified using a (nonlinear) autoregressive model with exogenous covariates (ARX).

If the exogenous covariate process $\{\boldsymbol{X}_t\}$ is stationary and ergodic, then there exists a stationary and ergodic solution $\{Y_t\}$ to the model (3) with the link function h given by (5) provided that the function r satisfies

$$|r(\boldsymbol{y}_1, \boldsymbol{\lambda}_1; \boldsymbol{\theta}_1) - r(\boldsymbol{y}_2, \boldsymbol{\lambda}_2; \boldsymbol{\theta}_1)| \leq \sum_{i=1}^{p} \alpha_i |y_{1i} - y_{2i}| + \sum_{j=1}^{q} \beta_j |\lambda_{1j} - \lambda_{2j}| \tag{6}$$

for all $\boldsymbol{y}_i = (y_{1i}, \ldots, y_{1p})^\top \in [0, \infty)^p$, $\boldsymbol{\lambda}_i = (\lambda_{i1}, \ldots, \lambda_{iq})^\top \in [0, \infty)^q$, $i = 1, 2$, for all $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}^{(1)}$ and for some $\alpha_i \geq 0$, $\beta_j \geq 0$, $i = 1, \ldots, p$, $j = 1, \ldots, q$ such that

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1, \tag{7}$$

see (Aknouche and Francq 2021).

Model (5) includes as a special case a linear model, which can be for non-negative components of $\boldsymbol{X}_t$ formulated as (3) for a link function

$$h(\boldsymbol{y}, \boldsymbol{l}, \boldsymbol{x}; \boldsymbol{\theta}) = \omega + \boldsymbol{\alpha}^\top \boldsymbol{y} + \boldsymbol{\beta}^\top \boldsymbol{l} + \boldsymbol{\pi}^\top \boldsymbol{x} \tag{8}$$

for $\boldsymbol{\theta} = (\omega, \boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\pi}^\top)^\top$ with $\omega > 0$, $\boldsymbol{\alpha} \in [0, 1]^p$, $\boldsymbol{\beta} \in [0, 1]^q$ and $\boldsymbol{\pi} \in (0, \infty)^m$. The condition for existence of a stationary and ergodic solution then reduces to (7). In that case

$$\mathbb{E}\,[Y_t|\mathcal{I}_{t-1}] = \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j} + \boldsymbol{\pi}^\top \boldsymbol{X}_{t-1}. \tag{9}$$

A model with the conditional mean specified by (9) and Poisson $\mathcal{F}_\lambda$, often referred to as PARX, has been proposed by Agosto et al. (2016). Various properties of this model and the corresponding maximum likelihood estimator of $\boldsymbol{\theta}$ are provided therein.

The PARX model often serves as a baseline model in practical applications, so a goodness-of-fit test for this particular model is of special importance in practice.

Remark 1 It is assumed after (3) that the vector $\boldsymbol{\theta}$ can be split into $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ such that in (5) the function r depends on endogenous variables and $\boldsymbol{\theta}_1$, while π depends on exogenous covariates $\boldsymbol{X}_{t-1}$ and $\boldsymbol{\theta}_2$. This division into $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, and r and π is not always unique. A typical example is a situation when h contains an intercept, because that can be included to both r and π. For the goodness–fit–test defined in the next section this non-uniqueness is not important.

Furthermore, note that if h contains an intercept, then the constant term can be treated as one of the components of the covariate process. To see that consider the linear model (9). For this model one can take $\boldsymbol{\theta}_1 = (\omega, \boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$ and $\boldsymbol{\theta}_2 = \boldsymbol{\pi}$ and exogenous covariates $\boldsymbol{X}_t$. Alternatively, it is possible to work with covariate process of vectors $\check{\boldsymbol{X}}_t = (1, \boldsymbol{X}_t^\top)^\top$ so that $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$ and $\boldsymbol{\theta}_2 = (\omega, \boldsymbol{\pi}^\top)^\top$. In that case the vector $\boldsymbol{Z}_t$ defined in (4) would contain an extra element 1. However, in the following we implicitly assume that we work with a non-trivial $\boldsymbol{X}_t$ (without an constant term) due to simpler dimensionality. This is also in agreement with the approach in Agosto et al. (2016).

3 Goodness-of-fit test

Let $\{(Y_t, \boldsymbol{X}_{t-1})\}_{t=1}^{T}$ be observed data. We develop test for the null hypothesis

$$\mathcal{H}_0 : \{(Y_t, \boldsymbol{X}_{t-1})\}_{t=1}^{T} \text{ follow model ((3)) with a specified class of distributions } \mathcal{F}_\lambda, \\ \text{and specified function } h(\cdot; \boldsymbol{\theta}_0) \text{ for some inner point } \boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{01}^\top, \boldsymbol{\theta}_{02}^\top)^\top \in \boldsymbol{\Theta}_0,$$

against a general alternative H_1 : not H_0, where $\boldsymbol{\Theta}_0$ stands for a compact subset of the parametric space $\boldsymbol{\Theta}$. The model orders (p, q) are implicitly specified by the function h, so they are assumed to be known under the null hypothesis.

Our test utilizes the characterization of Bierens (1982) which for a random variable U and a d-dimensional random vector $\boldsymbol{W}$ may be formulated via the following equivalence relation

$$\mathbb{E}[U|\boldsymbol{W}] = \mu_0(\boldsymbol{W}) \Longleftrightarrow \mathbb{E}[\{U - \mu_0(\boldsymbol{W})\}e^{\mathrm{i}\boldsymbol{v}^\top \boldsymbol{W}}] = 0, \ \forall \boldsymbol{v} \in \mathbb{R}^d, \tag{10}$$

where $\mathrm{i} = \sqrt{-1}$ and $\mu_0(\cdot)$ denotes a specific regression function possibly involving an unknown parameter vector.

In the current context, we rephrase (10) for model (3) in terms of the conditional probability generating function (PGF). Recall that a PGF of a count variable Y is defined as $g_Y(u) := \mathbb{E}(u^Y)$ for $u \in \mathbb{R}$ for which the expectation is finite. The PGF is always well defined and, as a power series, is absolutely convergent and differentiable for $|u| \leq 1$, and thus the probabilities $\mathsf{P}(Y = k), \ k = 0, 1, ...,$ may be obtained from the derivatives of $g_Y(u)$ at $u = 0$, so that the distribution of Y is uniquely determined from the values of $g_Y(\cdot)$ in a neighborhood of the origin; see Esquível (2008) and (Zayed 1996, Section 34.3.4). Let $g_\lambda(u)$ be the PGF corresponding to F_λ. Under $\mathcal{H}_0$ it holds that

$$\mathbb{E}[u^{Y_t}|I_{t-1}] = \mathbb{E}[u^{Y_t}|\boldsymbol{Z}_t] = g_{\lambda_t}(u) = g_{h(\boldsymbol{Z}_t;\boldsymbol{\theta}_0)}(u), \quad u \in [0, 1].$$

Hence, the characterization featuring in (10) can be reformulated that H_0 holds if and only if

$$\mathbb{E}[\{u^{Y_t} - g_{\lambda_t}(u)\}e^{\mathrm{i}\boldsymbol{v}^\top \boldsymbol{Z}_t}] = 0, \forall u \in (0, 1), \boldsymbol{v} \in \mathbb{R}^d, \quad \lambda_t = h(\boldsymbol{Z}_t; \boldsymbol{\theta}_0). \tag{11}$$

Our test procedure will be based on a weighted L_2 norm of an estimator of the expectation in (11). Let $\widehat{\boldsymbol{\theta}} := \widehat{\boldsymbol{\theta}}_T$ be a suitable estimator of the model parameter $\boldsymbol{\theta}$ and define

$$\widehat{\varepsilon}_t(u) = u^{Y_t} - g_{\widehat{\lambda}_t}(u), \ t = 1, ..., T, \tag{12}$$

where $\widehat{\lambda}_t$ are for $t = 1, \ldots, T$, defined recursively as $\widehat{\lambda}_t = h(\widehat{\boldsymbol{Z}}_t; \widehat{\boldsymbol{\theta}})$, with $\widehat{\boldsymbol{Z}}_t$ being an estimated version of $\boldsymbol{Z}_t$, computed from some initial values for Y_t, $\widehat{\lambda}_t$ for $t \leq 0$, i.e., $\widehat{\boldsymbol{Z}}_t = (Y_{t-1}, \ldots, Y_{t-p}, \widehat{\lambda}_{t-1}, \ldots, \widehat{\lambda}_{t-q}, \boldsymbol{X}_{t-1}^\top)^\top$ for $t \geq p+1$. Let $W : \mathbb{R}^{d+1} \to (0, \infty)$ be a non-negative weight function. Define the test statistic

$$\Delta_{T,W} = T \int_0^1 \int_{\mathbb{R}^d} |D_T(u; \boldsymbol{v})|^2 \, W(u, \boldsymbol{v}) \mathrm{d}\boldsymbol{v} \mathrm{d}u, \tag{13}$$

where

$$D_T(u; \boldsymbol{v}) = \frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_t(u) e^{\mathrm{i} \boldsymbol{v}^\top \widehat{\boldsymbol{Z}}_t}, \;\; (u, \boldsymbol{v}) \in (0, 1) \times \mathbb{R}^d$$

is an estimate of the expectation featuring in (11). The test statistic $\Delta_{T,W}$ in (13) is an L_2-type distance statistic which, at least for large sample size T, should be small under the null hypothesis H_0. On the other hand, it follows from the equivalence in (10) that $\mathbb{E}\,[\{u^{Y_t} - g_{\lambda_t}(u)\} e^{\mathrm{i} \boldsymbol{v}^\top \boldsymbol{Z}_t}] \neq 0$ whenever the conditional PGF of Y_t is not correctly specified, i.e., if either λ_t or $\mathcal{F}_\lambda$ is misspecified. This nonzero quantity is estimated by $D_T(u; \boldsymbol{v})$, which implies that the test statistic $\Delta_{T,W}$ is expected to be large under alternatives. Therefore large values of the test statistic indicate that (11) is violated and $\mathcal{H}_0$ should be rejected.

The estimator $\widehat{\boldsymbol{\theta}}$ can be either a maximum likelihood estimator (MLE), or, more generally, a Poisson quasimaximum likelihood estimator (QMLE). The latter method has been studied in Ahmad and Francq (2016) and Aknouche and Francq (2021), where it is shown that the resulting estimator is consistent and asymptotically normal under mild regularity conditions.

For practical computation of $\Delta_{T,W}$, it is useful to notice that

$$|D_T(u; \boldsymbol{v})|^2 = \frac{1}{T^2} \sum_{t,s=1}^{T} \widehat{\varepsilon}_t(u) \widehat{\varepsilon}_s(u) \cos\left(\boldsymbol{v}^\top (\widehat{\boldsymbol{Z}}_t - \widehat{\boldsymbol{Z}}_s)\right). \tag{14}$$

If the weight function can be decomposed as $W(u, \boldsymbol{v}) = w(u)\omega(\boldsymbol{v})$, with $w(\cdot)$ and $\omega(\cdot)$ satisfying some additional assumptions, then further simplification is possible; see Sect. 5 for more details on the practical aspects of computing the test statistic $\Delta_{T,W}$. Furthermore, it is mentioned in the previous sections that the linear Poisson model (PARX) has been used in various applications, so the computation of $\Delta_{T,W}$ for this class of models is explored in more detail Sect. 5.3.

Remark 2 Note that the integral in (13) is computed for $u \in [0, 1]$, which is an approach consistent with most previous works, see, e.g., Rueda et al. (1991); Baringhaus and Henze (1992); Novoa-Muñoz and Jiménez-Gamero (2014), Nakamura and Pérez-Abreu (1993), and Jiménez-Gamero and Batsidis (2017), among others. In this connection, the main asymptotic analysis is due to Feuerverger (1988) and Rémillard and Theodorescu (2000) showing weak convergence of the empirical PGF process on [0,1], while Szűcs (2005) also studies the bootstrap in the context of the empirical

PGF. In principle though, integration can be carried over the interval $[-1, 1]$, or over an interval $I \subset [-1, 1]$ containing the origin, and the asymptotic distribution of the corresponding test statistics should be obtained analogously as for $\Delta_{T,W}$, but here we keep in line with most authors and work on [0,1].

Remark 3 The test statistic $\Delta_{T,W}$ is designed to be sensitive to all violations from $\mathcal{H}_0$, so it not only detects an incorrect specification of the conditional mean, but also violation of the distributional assumption specified by $\mathcal{F}_\lambda$. If one is interested solely in the specification of $\mathbb{E}[Y_t|\mathcal{I}_{t-1}]$, then a suitable test can be derived from (10) formulated directly for Y_t as it is done in Bierens (1982) for a general regression situation. Indeed, $\mathbb{E}[Y_t|\mathcal{I}_{t-1}] = h(\mathbf{Z}_t; \boldsymbol{\theta}_0)$ if and only if $\mathbb{E}[\{Y_t - h(\mathbf{Z}_t; \boldsymbol{\theta}_0)\} e^{\mathrm{i}\boldsymbol{v}^\top \mathbf{Z}_t}] = 0$, for all $\boldsymbol{v} \in \mathbb{R}^d$, so a suitable test statistic is obtained by replacing $\widehat{\varepsilon}_t(u)$ in (13) by $\widehat{\eta}_t = Y_t - h(\widehat{\mathbf{Z}}_t; \widehat{\boldsymbol{\theta}})$.

Remark 4 An important advantage of $\Delta_{T,W}$ defined in (13) is that it is derived from the *equivalence relationship* in (11). Relaxing this equivalence can lead to a construction of simpler test statistics. See that $\mathcal{H}_0$ implies that $\{\varepsilon_t(u)\}$, with $\varepsilon_t(u) := u^{Y_t} - g_{\lambda_t}(u)$, is a martingale difference sequence for all $u \in [0, 1]$. Define

$$\Delta^{(0)}_{T,w} = T \int_0^1 \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_t(u) \right)^2 w(u) \mathrm{d}u, \qquad \Delta^{(1)}_{T} = \sqrt{T} \sup_{0<u<1} \left| \frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_t(u) \right|$$

for some weight function $w : [0, 1] \to (0, \infty)$. Then $\Delta^{(0)}_{T,w}$ and $\Delta^{(1)}_{T}$ are the Cramér von-Mises-type and the Kolmogorov-Smirnov-type test statistics, respectively, derived from $D_T(u; \mathbf{0})$. Hence, they are expected to be small under $\mathcal{H}_0$ and they should, to some extend, react to violations from $\mathcal{H}_0$. The small sample behavior of these two test statistics is compared to $\Delta_{T,W}$ in a Monte Carlo study in Sect. 6.

Remark 5 The covariate process $\{\boldsymbol{X}_t\}$ is treated as stochastic, and the asymptotic distribution of the test statistic $\Delta_{T,W}$ is derived in Sect. 4 under the assumption that it is stationary and ergodic. This approach is consistent with Agosto et al. (2016) or Aknouche and Francq (2021).

However, a large number of count time series encountered in practice contain a deterministic periodic component. For instance, daily number of accidents exhibit weak periodicity, monthly series of tourists counts show month periodicity, or daily counts of respiratory diseases can exhibit even more complicated periodic behavior; see, e.g., Prezotti Filho et al. (2021). If there are no other regressors next to the purely deterministic periodicity, then (3) fits into the class of periodic count time series models considered by Bentarzi and Aries (2022). In this case the limiting distribution of $\Delta_{T,W}$ could be derived along the same lines as in Sect. 4 under the assumption that the process $\{Y_t\}$ is strictly periodically stationary and periodically ergodic, see, e.g., Aknouche and Al-Eid (2012) for formal definitions. An estimator of the model parameters can be constructed using the Poisson QML (Bentarzi and Aries 2022).

Kedem and Fokianos (2002) deal with regression models for count time series and consider log-linear type of models with "well-behaved" regressors, see Assumption A4 in their Section 1.4 for an exact formulation and examples in their Chapter 4. In particular, their approach allows a mix of periodic deterministic variables and stationary stochastic variables. Hence, it is expected that the limiting distribution of $\Delta_{T,W}$ could be derived similarly under more general assumptions, when part of the vector is deterministic periodic function and the rest forms stationary ergodic series. Such covariate process is considered in the practical data application in Sect. 7.

4 Asymptotic distribution

This section describes the asymptotic behavior of the test statistic $\Delta_{T,W}$ under the null hypothesis (Sect. 4.1) as well as under some alternatives (Sect. 4.3). Since some of the assumptions can be substantially weakened for an autoregressive type of model, where $q = 0$, the behavior under the null for these models is treated in more detail in Sect. 4.2.

4.1 Behavior under the null hypothesis

We start with some assumptions on the link function and the weight function. We use the notation $\mathsf{int}(A)$ for the interior of a set A and $\|\boldsymbol{x}\|$ is the Euclidean norm of a vector $\boldsymbol{x}$.

Assumption $\mathcal{A}1$
Let $h : [0,\infty)^{p+q} \times \mathbb{R}^{m+K} \to (c_h, \infty)$ for some $c_h > 0$ be a function of the form (5) for some continuous non-negative functions $r : [0,\infty)^{p+q} \times \mathbb{R}^{K_1} \to (0,\infty)$ and $\pi : \mathbb{R}^{m+K_2} \to (0,\infty)$. Let $\boldsymbol{\Theta} = \boldsymbol{\Theta}^{(1)} \times \boldsymbol{\Theta}^{(2)} \subset \mathbb{R}^K$ be such that (6) holds for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for some $\alpha_i, \beta_j, i = 1,\ldots,p, j = 1,\ldots,q$, satisfying (7).

Assumption $\mathcal{A}2$
Let $\{\boldsymbol{X}_t\}$ be a strictly stationary and ergodic sequence of m-dimensional random vectors. Let $\{\mathcal{F}_\lambda\}$ satisfy (2) and let $\{Y_t\}$ be the stationary and ergodic solution of (3) with $\lambda_t = h(\boldsymbol{Z}_t; \boldsymbol{\theta}_0)$, where $\boldsymbol{Z}_t$ is defined by (4), for a true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{01}^\top, \boldsymbol{\theta}_{02}^\top)^\top \in \mathsf{int}(\boldsymbol{\Theta}_0)$, where $\boldsymbol{\Theta}_0$ is a compact subset of $\boldsymbol{\Theta}$. Assume further that $\mathbb{E}\,\|\boldsymbol{Z}_t\|^2 < \infty$.

Assumption $\mathcal{A}3$
Let r and π be the functions from Assumption $\mathcal{A}1$. Let $r(\boldsymbol{y},\cdot,\cdot)$ be three times continuously differentiable on $(0,\infty)^q \times \boldsymbol{\Theta}^{(1)}$ for all $\boldsymbol{y} \in (0,\infty)^p$ and denote

$$r_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\theta}_1) = \frac{\partial r(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\theta}_1)}{\partial \boldsymbol{\lambda}}, \quad r_{\boldsymbol{\theta}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\theta}_1) = \frac{\partial r(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}$$

for $\boldsymbol{\lambda} \in \mathbb{R}^q$ and $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}^{(1)}$. Suppose $r_{\boldsymbol{\lambda}}(\boldsymbol{y},\boldsymbol{\lambda};\boldsymbol{\theta}_1) = d(\boldsymbol{\theta}_1)$ is constant with respect to $\boldsymbol{y}$ and $\boldsymbol{\lambda}$ such that

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \rho(\boldsymbol{D}) < 1, \quad \boldsymbol{D} = \begin{pmatrix} & d(\boldsymbol{\theta}_1) \\ \boldsymbol{I}_{q-1} & \boldsymbol{0} \end{pmatrix},$$

where $\rho(\cdot)$ is a spectral radius and $\boldsymbol{I}_k$ is a $k \times k$ identity matrix. Let π be two times continuously differentiable with respect to $\boldsymbol{\theta}_2$ and denote

$$\pi_{\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta}_2) = \frac{\partial \pi(\boldsymbol{x};\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2}.$$

Assume that for the stationary variables $Y_t, \lambda_t, \boldsymbol{X}_t$ there exists $\kappa > 0$ such that

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta_0} \|\pi_{\boldsymbol{\theta}}(\boldsymbol{X}_{t-1};\boldsymbol{\theta}_2)\|^{\kappa} < \infty, \quad \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta_0} \|r_{\boldsymbol{\theta}}(\boldsymbol{Y}_{t-1:t-p}, \boldsymbol{\Lambda}_{t-1:t-q};\boldsymbol{\theta}_1)\|^{\kappa} < \infty, \tag{15}$$

where $\boldsymbol{Y}_{t-1:t-p} = (Y_{t-1}, \ldots, Y_{t-p})^\top$ and $\boldsymbol{\Lambda}_{t-1:t-q} = (\lambda_{t-1}, \ldots, \lambda_{t-q})^\top$.

Assumption $\mathcal{A}4$ Assume that the PGF $g_\lambda(u)$ of the distribution $\mathcal{F}_\lambda$, considered as a function $g_z(u)$ of $(u,z)^\top$, is twice continuously differentiable and that

$$\left|\frac{\partial g_z(u)}{\partial z}\right| \leq M_1, \quad \left|\frac{\partial^2 g_z(u)}{\partial z^2}\right| \leq M_2, \quad \left|\frac{\partial^2 g_z(u)}{\partial z \partial u}\right| \leq H(z)$$

for all $(u,z)^\top \in (0,1)\times(0,\infty)$, and some constants $M_1, M_2 < \infty$, where $H(\cdot)$ denotes a finite positive function.

Assumption $\mathcal{A}5$ Let $(Y_t, \lambda_t(\boldsymbol{\theta}))$ be the stationary solution of (3) for parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Let the second derivative $\frac{\partial^2 \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}$ exist and be continuous in $\boldsymbol{\theta}$ and let there exist a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^\top \quad \text{and} \quad \mathbb{E} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \frac{\partial^2 \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} \tag{16}$$

are finite. Furthermore, assume that $\mathbb{E}\, H^2(\lambda_t(\boldsymbol{\theta}_0)) < \infty$, where the function $H(\cdot)$ is from Assumption $\mathcal{A}4$.

Assumption $\mathcal{A}6$ Let $\widehat{\boldsymbol{\theta}}_T$ be an estimator of $\boldsymbol{\theta}$ satisfying

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \boldsymbol{s}_t(\boldsymbol{\theta}_0, Y_t, \mathcal{I}_{t-1}) + o_{\mathsf{P}}(1),$$

where $\boldsymbol{\theta}_0$ is from $\mathcal{A}2$, $\{\boldsymbol{s}_t(\boldsymbol{\theta}_0, Y_t, \mathcal{I}_{t-1})\}$ is a stationary martingale difference sequence with respect to the filtration $\{\mathcal{I}_t\}$ with a finite variance such that $T^{-1/2}\sum_{t=1}^{T} \boldsymbol{s}_t(\boldsymbol{\theta}_0, Y_t, \mathcal{I}_{t-1}) \xrightarrow{D} \mathsf{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ for some $K\times K$ positive semidefinite matrix $\boldsymbol{\Sigma}$.

Assumption $\mathcal{A}7$
Let $W : \mathbb{R}^{d+1} \to (0,1)$ be $W(u, \boldsymbol{v}) = w(u)\omega(\boldsymbol{v})$ for some measurable functions $w : (0,1) \to (0,\infty)$ and $\omega : \mathbb{R}^d \to (0,\infty)$ such that $\omega(\boldsymbol{v}) = \omega(-\boldsymbol{v})$ for all $\boldsymbol{v} \in \mathbb{R}^d$

and

$$0 < \int_0^1 w(u)\mathrm{d}u < \infty, \quad 0 < \int_{\mathbb{R}^d} ||\boldsymbol{v}||^k \omega(\boldsymbol{v})\mathrm{d}\boldsymbol{v} < \infty$$

for $k \leq 4$.

The existence of a stationary ergodic solution from Assumption $\mathcal{A}2$ is proved for h satisfying Assumption $\mathcal{A}1$ in (Aknouche and Francq 2021, Theorem 3.3). Under the same conditions, the process $(Y_t, \boldsymbol{X}_t^\top)^\top$ is stationary and ergodic and the same holds for the multivariate process $\{\boldsymbol{Z}_t\}$. Sufficient conditions for the existence of moments for the linear case with $p = q = 1$ are provided in (Aknouche and Francq 2021, Theorem 3.2). Also notice that we assume $\boldsymbol{\theta}_0$ lies in the interior of the parameter space $\boldsymbol{\Theta}_0$; for estimation and specification testing which allow the parameter to lie on the boundary of the parameter space we refer to Francq and Zakoïan (2007); Ahmad and Francq (2016); Cavaliere et al. (2024) and a discussion in Agosto et al. (2016).

Assumption $\mathcal{A}3$ is related to the smoothness of the mean function and the required conditions ensure that the effect of the initial values is asymptotically negligible. A similar set of assumptions is invoked by Aknouche and Francq (2021) to prove the asymptotic normality of the QMLE. Notice that Assumption $\mathcal{A}3$ is satisfied by a linear function (8) provided that (7) and $\mathcal{A}2$ hold.

Assumption $\mathcal{A}4$ ensures that the PGF g_λ as a function λ is sufficiently smooth. For a Poisson distribution, one can take $M_1 = M_2 = 1$ and $H(z) = z + 1$. Assumption $\mathcal{A}6$ is satisfied under some additional regularity conditions by the MLE, see Agosto et al. (2016) for a linear Poisson model. More generally, the estimator $\widehat{\boldsymbol{\theta}}_T$ may be the Poisson QMLE, see Ahmad and Francq (2016); Aknouche and Francq (2021).

Example 1 Since the PARX model (9) with Poisson $\mathcal{F}_\lambda$ is an important special case, we discuss the assumptions $\mathcal{A}1$–$\mathcal{A}6$ for this situation in more detail. Let $\{\boldsymbol{X}_t\}$ be a strictly stationary, ergodic and *non-negative* such that $\mathbb{E}\,\|\boldsymbol{X}_t\|^2 < \infty$. It follows from the above discussion that Assumption $\mathcal{A}4$ holds for the Poisson distribution, and the remaining assumptions simplify as follows:

- $\mathcal{A}1$ is satisfied if the parameters α_i and β_j, $i = 1, \ldots, p$, $j = 1, \ldots, q$, satisfy (7) and the parametric space $\boldsymbol{\Theta}$ is such that $\omega > c_h > 0$ for some c_h for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.
- $\mathcal{A}2$ reduces to the standard requirement that the true parameter is an inner point of a compact set and that $\mathbb{E}\,\lambda_t^2 < \infty$. Sufficient conditions for the latter are given in Agosto et al. (2016) for Markovian $\{\boldsymbol{X}_t\}$.
- $\mathcal{A}3$: Since r and π are both linear functions, $\mathcal{A}3$ is entailed by the previous assumptions.
- $\mathcal{A}5$: This assumption follows from the linearity of h and from the finiteness of the second-order moments of $\boldsymbol{Z}_t$.
- $\mathcal{A}6$: For $\mathcal{F}_\lambda$ Poisson, the two mentioned estimation methods, MLE and Poisson QMLE, coincide. Next to the previously stated assumptions, the following identifiability condition is required: $[\lambda_1(\boldsymbol{\theta}_A) = \lambda_1(\boldsymbol{\theta}_B)$ a.s. holds if and only if $\boldsymbol{\theta}_A = \boldsymbol{\theta}_B]$. A necessary condition for this is that the components of $\boldsymbol{X}_t$ are linearly independent. A detailed discussion on this condition is provided in (Aknouche and Francq 2021, page 261).

Define for $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^d$ and $u \in [0, 1]$

$$\alpha(\boldsymbol{v}_1, \boldsymbol{v}_2) = \cos\left(\boldsymbol{v}_1^\top \boldsymbol{v}_2\right) + \sin\left(\boldsymbol{v}_1^\top \boldsymbol{v}_2\right),$$
$$\boldsymbol{\beta}(u, \boldsymbol{v}_1) = -\mathbb{E}\left\{\frac{\partial g_z(u)}{\partial z}\Big|_{z=\lambda_1} \frac{\partial \lambda_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \alpha(\boldsymbol{Z}_1, \boldsymbol{v}_1)\right\}.$$

Note that $\boldsymbol{\beta}(u, \boldsymbol{v}_1)$ is finite for all $u \in [0, 1]$ and $\boldsymbol{v} \in \mathbb{R}^d$ due to Assumptions $\mathcal{A}4$ and $\mathcal{A}5$. The following theorem describes the asymptotic distribution of the proposed test statistic under the null hypothesis H_0.

Theorem 1 *Let Assumptions $\mathcal{A}1$–$\mathcal{A}7$ be satisfied. Then the test statistic $\Delta_{T,W}$ converges in distribution for $T \to \infty$ to $\int_0^1 \int_{\mathbb{R}^d} |\mathcal{Z}(u, \boldsymbol{v})|^2 W(u, \boldsymbol{v}) \mathrm{d}\boldsymbol{v}\mathrm{d}u$, where $\mathcal{Z} = \{\mathcal{Z}(u, \boldsymbol{v}), u \in [0, 1], \boldsymbol{v} \in \mathbb{R}^d\}$ is a centered Gaussian process with a covariance function*

$$\begin{aligned}\mathbb{E}\{\mathcal{Z}(u_1, \boldsymbol{v}_1)\,\mathcal{Z}(u_2, \boldsymbol{v}_2)\} =& \mathbb{E}\left[\left(u_1^{Y_1} - g_{\lambda_1}(u_1)\right)\alpha(\boldsymbol{Z}_1, \boldsymbol{v}_1) + \boldsymbol{s}_1(\boldsymbol{\theta}_0, Y_1, \mathcal{I}_0)^\top \boldsymbol{\beta}(u_1, \boldsymbol{v}_1)\right] \\ &\times \left[\left(u_2^{Y_1} - g_{\lambda_1}(u_2)\right)\alpha(\boldsymbol{Z}_1, \boldsymbol{v}_2) + \boldsymbol{s}_1(\boldsymbol{\theta}_0, Y_1, \mathcal{I}_0)^\top \boldsymbol{\beta}(u_2, \boldsymbol{v}_2)\right].\end{aligned}$$

The assertion provides an approximation of the distribution of $\Delta_{T,W}$ under the null hypothesis. It follows from Theorem 1 that $\Delta_{T,W}$ is asymptotically distributed as an infinite weighted sum of χ_1^2 distributed random variables, where, however, the weights depend on the unknown parameters via a highly non-trivial way. Hence, it is extremely complicated to use this asymptotic distribution in order to determine critical points and actually carry out the test. Instead, in Sect. 5 we recommend the use of a bootstrap test which is straightforward to apply.

4.2 Behavior under the null for an ARX type of model

An important special case of the model (3) is obtained for $q = 0$, when the conditional mean λ_t depends only on the past values of Y_t and on $\boldsymbol{X}_{t-1}$ via some kind of an ARX structure. Some of the assumptions specified in the previous section can be weakened for these models, so we treat this situation in more detail. If $q = 0$, then

$$h(\boldsymbol{y}, \boldsymbol{x}; \boldsymbol{\theta}) = r(\boldsymbol{y}; \boldsymbol{\theta}_1) + \pi(\boldsymbol{x}; \boldsymbol{\theta}_2), \quad \boldsymbol{y} \in [0, \infty)^p, \boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top \in \mathbb{R}^K$$

for some functions r and π, and $\boldsymbol{Z}_t = (Y_{t-1}, \ldots, Y_{t-q}, \boldsymbol{X}_{t-1}^\top)^\top$ depends solely on past values of Y_t and on $\boldsymbol{X}_{t-1}$. This fact enables us to reduce and simplify the assumptions required in $\mathcal{A}3$, $\mathcal{A}5$ and $\mathcal{A}7$, into the following versions:

Assumption $\mathcal{A}3^*$

Let $r : [0, \infty)^p \times \mathbb{R}^{K_1} \to (0, \infty)$ and $\pi : \mathbb{R}^{K_2+m} \to (0, \infty)$ be functions such that $r(\boldsymbol{y}; \cdot)$ and $\pi(\boldsymbol{x}; \cdot)$ are twice continuously differentiable for all $\boldsymbol{y} \in [0, \infty)^p$ and $\boldsymbol{x} \in \mathbb{R}^m$. Assume that there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\partial r(\boldsymbol{Y}_{t-1:t-p};\boldsymbol{\theta}_1)}{\partial \theta_i} \frac{\partial r(\boldsymbol{Y}_{t-1:t-p};\boldsymbol{\theta}_1)}{\partial \theta_j} \right| < \infty, \quad \mathbb{E} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\partial^2 r(\boldsymbol{Y}_{t-1:t-p};\boldsymbol{\theta}_1)}{\partial \theta_i \partial \theta_j} \right| < \infty,$$

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\partial \pi(\boldsymbol{X}_{t-1};\boldsymbol{\theta}_2)}{\partial \theta_k} \frac{\partial \pi(\boldsymbol{X}_{t-1};\boldsymbol{\theta}_2)}{\partial \theta_l} \right| < \infty, \quad \mathbb{E} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left| \frac{\partial^2 \pi(\boldsymbol{X}_{t-1};\boldsymbol{\theta}_2)}{\partial \theta_k \partial \theta_l} \right| < \infty$$

for all $i, j = 1, \ldots, K_1, k, l = K_1 + 1, \ldots, K$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^\top$, $\boldsymbol{Y}_{t-1:t-p}$ is defined below (15) and Y_t, $\boldsymbol{X}_t$ are the stationary variables from Assumption $\mathcal{A}2$.

Assumption $\mathcal{A}5^*$ Assume that $\mathbb{E}\, H^2(\lambda_t) < \infty$, where the function $H(\cdot)$ is from Assumption $\mathcal{A}4$ and (Y_t, λ_t) are the stationary variables from Assumption $\mathcal{A}2$.

Assumption $\mathcal{A}7^*$ Let $W : \mathbb{R}^{d+1} \to (0, 1)$ be $W(u, \boldsymbol{v}) = w(u)\omega(\boldsymbol{v})$ for some measurable functions $w : (0, 1) \to (0, \infty)$ and $\omega : \mathbb{R}^d \to (0, \infty)$ such that $\omega(\boldsymbol{v}) = \omega(-\boldsymbol{v})$ for all $\boldsymbol{v} \in \mathbb{R}^d$ and

$$0 < \int_0^1 w(u)\mathrm{d}u < \infty, \quad 0 < \int_{\mathbb{R}^d} \omega(\boldsymbol{v})\mathrm{d}\boldsymbol{v} < \infty.$$

Assumption $\mathcal{A}7^*$ is clearly a weakened version of $\mathcal{A}7$, because only integrability of ω is required. Assumptions $\mathcal{A}3^*$ and $\mathcal{A}5^*$ together imply $\mathcal{A}5$ and the main advantage of this reformulation is that conditions in $\mathcal{A}3^*$ are very easily verifiable in practical applications. For instance, if both r and π are linear, then the conditions in $\mathcal{A}3^*$ reduce to the requirement of finiteness of second-order moments of Y_t and $\boldsymbol{X}_t$.

Theorem 2 *Let assumptions $\mathcal{A}1$–$\mathcal{A}2$ hold for $q = 0$, and $\mathcal{A}3^*$, $\mathcal{A}4$, $\mathcal{A}5^*$, $\mathcal{A}6$ and $\mathcal{A}7^*$ be satisfied. Then the test statistic $\Delta_{T,W}$ converges in distribution for $T \to \infty$ to a random variable $\int_0^1 \int_{\mathbb{R}^d} |\mathcal{Z}(u, \boldsymbol{v})|^2 W(u, \boldsymbol{v})\mathrm{d}\boldsymbol{v}\mathrm{d}u$, where $\mathcal{Z}$ is the process specified in Theorem 1.*

4.3 Behavior under some alternatives

The limit behavior of the test statistic is provided also for some types of alternatives. To this end analogously to the notation below (15), we write $\boldsymbol{Y}_{k:l}$ for $(Y_k, Y_{k-1}, \ldots, Y_{k-(k-l)})^\top$ and $\boldsymbol{\Lambda}_{k:l}$ for $(\lambda_k, \lambda_{k-1}, \ldots, \lambda_{k-(k-l)})^\top$, for any integers (k, l) with $k \geq l$.

Let $p_0 > 0$, $q_0 \geq 0$, $K_0 \geq 1$ and write $h_0 : [0, \infty)^{p_0+q_0} \times \mathbb{R}^{m+K_0} \to (0, \infty)$ for a particular link function. Consider the null hypothesis $\mathcal{H}_0$ whereby the link function is specified by h_0 and a fixed family of distributions $\{\mathcal{F}_\lambda, \lambda > 0\}$, that is

$$\begin{aligned} \mathcal{H}_0 : \text{there exists } \boldsymbol{\theta} \in \mathsf{int}(\boldsymbol{\Theta}_0) \text{ such that } Y_t|\mathcal{I}_{t-1} \sim \mathcal{F}_{\lambda_t} \\ \lambda_t = h_0(\boldsymbol{Y}_{t-1:t-p_0}, \boldsymbol{\Lambda}_{t-1:t-q_0}, \boldsymbol{X}_{t-1}; \boldsymbol{\theta}), \end{aligned}$$

for $\boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$ a compact set. A general alternative admits an incorrect specification of the conditional mean function as well as the conditional distribution. Specifically suppose that the actual model is given by

$$Y_t|\mathcal{I}_{t-1} \sim \mathcal{F}^A_{\lambda_t} \quad \lambda_t = h_A(\boldsymbol{Y}_{t-1:t-p_A}, \boldsymbol{\Lambda}_{t-1:t-q_A}, \boldsymbol{X}_{t-1}; \boldsymbol{\theta}_A)$$

for some family of distributions $\{\mathcal{F}^A_\lambda, \lambda > 0\}$, a link function $h_A : [0,\infty)^{p_A+q_A} \times \mathbb{R}^{m+K_A} \to (0,\infty)$, with $p_A > 0, q_A \geq 0, K_A \geq 1$, and for some $\boldsymbol{\theta}_A \in \mathbb{R}^{K_A}$. Hence if the true pair $(\mathcal{F}^A_\lambda, h^A)$ differs from $(\mathcal{F}_\lambda, h_0)$ then the model is not correctly specified under $\mathcal{H}_0$. Let $\Delta_{T,W}$ be the test statistic defined by (13) for testing $\mathcal{H}_0$, based on some estimator $\widehat{\boldsymbol{\theta}}_T$ computed under $\mathcal{H}_0$, with the sequence $\{\widehat{\lambda}_{t,0}\}$ defined recursively using the link function h_0, and a d-dimensional sequence $\{\widehat{\boldsymbol{Z}}_{t,0}\}$ where $d = p_0 + q_0 + m$. Assume that the PGF $g_\lambda(u)$ of $\mathcal{F}_\lambda$ satisfies

$$|\partial g_\lambda(u)/\partial\lambda| < M_1 \text{ for all } u \in [0,1] \text{ and all } \lambda \in (0,\infty) \tag{17}$$

for some $M_1 > 0$ and that there exists $\boldsymbol{\theta}_0 \in \mathrm{int}(\boldsymbol{\Theta}_0)$ such that

$$\widehat{\boldsymbol{\theta}}_T \xrightarrow{\mathsf{P}} \boldsymbol{\theta}_0, \quad T \to \infty. \tag{18}$$

Notice that $\boldsymbol{\theta}_0$ generally differs from the true parameter $\boldsymbol{\theta}_A$. If the mean function h is specified correctly, i.e., if $h_A = h_0$, then it is possible to construct $\widehat{\boldsymbol{\theta}}$ such that (18) is satisfied for $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_A$. For instance, a Poisson QML estimator $\widehat{\boldsymbol{\theta}}$ has this property if $\mathcal{F}^A_\lambda$ satisfies (2) and some other standard assumptions, see Aknouche and Francq (2021). On the other hand, if $h_A \neq h_0$, then $\boldsymbol{\theta}_0$ is typically a point which optimizes the corresponding theoretical counterpart of the estimation criterion, see, e.g., White (1982) for the ML case. In the i.i.d. setting assumptions as that in (18) are rather common, we refer e.g., to eqn. (2.7) of Hušková and Meintanis (2010) and Theorem 5.1 in Meintanis et al. (2014), but also see Section 3 of Jiménez-Gamero et al. (2020) and Theorem 5 of Jiménez-Gamero (2014) for similar assumptions in analogous situations.

In the following two sections we discuss separately the two important situations: an incorrectly specified conditional distribution combined with the correct link function, i.e., $(\mathcal{F}^A_\lambda, h_A) = (\mathcal{F}^A_\lambda, h_0)$, $\mathcal{F}^A_\lambda \neq \mathcal{F}_\lambda$, and a correct distribution $\mathcal{F}^A_\lambda = \mathcal{F}_\lambda$ combined with an incorrect mean function $h_A \neq h_0$.

4.3.1 Misspecified conditional distribution

Consider first an alternative whereby the mean function is correctly specified $h_A = h_0$, but $\mathcal{F}_\lambda \neq \mathcal{F}^A_\lambda$, and write g^A_λ for the PGF corresponding to $\mathcal{F}^A_\lambda$. As mentioned above, in this case it is possible to assume that (18) holds with $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_A$.

Theorem 3 *Let $\{Y_t, \lambda_t\}$ and $\{\boldsymbol{X}_t\}$ satisfy $\mathcal{A}2$ with the conditional distribution $\mathcal{F}^A_\lambda$, the link function h_0 and a true parameter $\boldsymbol{\theta}_A$. Let $\mathcal{A}1$, $\mathcal{A}3$, $\mathcal{A}5$, $\mathcal{A}7$, (17) and (18) hold for $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_A$. Then*

$$\frac{\Delta_{T,W}}{T} \xrightarrow{\mathsf{P}} \int_0^1 \int_{\mathbb{R}^d} \zeta^2(u, \boldsymbol{v}) W(u, \boldsymbol{v}) \mathrm{d}\boldsymbol{v}\mathrm{d}u,$$

as $T \to \infty$, *where*

$$\zeta(u, \boldsymbol{v}) = \mathbb{E}\left\{\left[g^A_{\lambda_1}(u) - g_{\lambda_1}(u)\right]\alpha(\boldsymbol{Z}_1, \boldsymbol{v})\right\}, \quad \boldsymbol{Z}_t = (\boldsymbol{Y}^\top_{t-1:t-p_0}, \boldsymbol{\Lambda}^\top_{t-1:t-q_0}, \boldsymbol{X}^\top_{t-1}).$$

4.3.2 Misspecified conditional mean

Consider now a situation whereby the conditional distribution is correctly specified under the null hypothesis, but the link function is misspecified, i.e., $\mathcal{F}^A_\lambda = \mathcal{F}_\lambda$ but $h_A \neq h_0$. This class of alternatives involves the following important examples:

- A model with incorrect orders. As an example consider a test for ARX-type linear model (9) with $p = 1, q = 0$, while the data are generated from (9) with $p > 1$ or $q > 0$.
- Test for a linear model, while the true model is nonlinear. As a particular example consider a test of (9) with $p = q = 1$, while the true model is log-linear with covariates

$$\log(\lambda_t) = \omega + \alpha_1 \log(Y_{t-1} + 1) + \beta_1 \log \lambda_{t-1} + \boldsymbol{\pi}^\top \boldsymbol{X}_{t-1}.$$

  Note that such log-linear Poisson models without covariates are studied in Fokianos and Tjøstheim (2011).
- A model with incorrectly specified π function which models the dependence on the exogenous covariates, see Sect. 6 for particular examples.

We provide the asymptotic behavior of $\Delta_{T,W}$ in two practically important situations, when either h_0 corresponds to a linear model, or when h_0 is nonlinear with $q_0 = 0$, i.e., the tested link h_0 corresponds to an ARX type of a model. For both situations we show that the test based on $\Delta_{T,W}$ is consistent.

Let us start with the latter (nonlinear) case, whereby we test the link function of form $h_0(\boldsymbol{y}, \boldsymbol{x}; \boldsymbol{\theta}) = r_0(\boldsymbol{y}; \boldsymbol{\theta}_1) + \pi_0(\boldsymbol{x}; \boldsymbol{\theta}_2)$ for some $r : [0, \infty)^{p_0} \times \mathbb{R}^{K_1} \to (0, \infty)$ and $\pi : \mathbb{R}^{K_2+m} \to (0, \infty)$.

Similarly as in Sect. 4.2, we have $\widehat{\boldsymbol{Z}}_{t,0} = \boldsymbol{Z}_{t,0} = (\boldsymbol{Y}^\top_{t-1:t-p_0}, \boldsymbol{X}^\top_{t-1})^\top$ for $t \geq p_0+1$. Notice that nothing is assumed about the true orders (p_A, q_A) of the true link function h_A, so these can be completely arbitrary. In the following we denote $\lambda_{t,A} = \mathbb{E}\,[Y_t|\mathcal{I}_{t-1}]$ the true conditional mean satisfying $\lambda_{t,A} = h_A(\boldsymbol{Y}_{t-1:t-p_A}, \boldsymbol{\Lambda}_{t-1:t-q_A}, \boldsymbol{X}_{t-1}; \boldsymbol{\theta}_A)$.

Theorem 4 *Let* $h_0 : [0, \infty)^{p_0} \times \mathbb{R}^{K_0+m} \to (0, \infty)$ *be a function satisfying* $\mathcal{A}1$ *on* $\boldsymbol{\Theta} \subset \mathbb{R}^{K_0}$ *for some* $p_0 > 0$, $K_0 \geq 1$. *Let* $\{(Y_t, \lambda_{t,A})\}$ *and* $\{\boldsymbol{X}_t\}$ *satisfy* $\mathcal{A}2$ *with a link function* h_A, *parameter* $\boldsymbol{\theta}_A$ *and distribution* $\mathcal{F}_\lambda$. *Assume that* (17), (18) *and* $\mathcal{A}7^*$ *hold, and there exists a neighborhood* $\mathcal{V}(\boldsymbol{\theta}_0)$ *of* $\boldsymbol{\theta}_0$ *such that*

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left|h_0(\boldsymbol{Y}_{t-1:t-p_0}, \boldsymbol{X}_{t-1}; \boldsymbol{\theta})\right| < \infty, \quad \mathbb{E} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \left|\frac{\partial h_0(\boldsymbol{Y}_{t-1:t-p_0}, \boldsymbol{X}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right| < \infty.$$

Then

$$\frac{1}{T}\Delta_{T,W} \xrightarrow{\mathsf{P}} \int_0^1 \int_{\mathbb{R}^d} \zeta^2(u, \boldsymbol{v}) w(u) \omega(\boldsymbol{v}) \mathrm{d}\boldsymbol{v} \mathrm{d}u$$

as $T \to \infty$, *where*

$$\zeta(u,\boldsymbol{v}) = \mathbb{E}\left\{\left[g_{\lambda_{1,A}}(u) - g_{\lambda_{1,0}}(u)\right]\alpha(\boldsymbol{Z}_{1,0},\boldsymbol{v})\right\}, \quad \lambda_{1,0} = h_0(\boldsymbol{Y}_{0:1-p_0}, \boldsymbol{X}_0; \boldsymbol{\theta}_0).$$

Since the linear model (9) is a very important special case, we consider properties of the goodness-of-fit test for a null hypothesis with a linear h_0 in more detail. For simplicity, we focus on a linear model with orders $p = q = 1$, that is

$$h_0(y,\lambda,\boldsymbol{x},;\boldsymbol{\theta}) = \omega + \alpha y + \beta\lambda + \boldsymbol{\gamma}^\top \boldsymbol{x} \tag{19}$$

for $\boldsymbol{\theta} = (\omega,\alpha,\beta,\boldsymbol{\gamma}^\top)^\top$. Let $\boldsymbol{\Theta}$ be a subset of $\mathbb{R}^{3+m}$ such that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{\theta} = (\omega,\alpha,\beta,\boldsymbol{\gamma}^\top)^\top$, it holds that $\omega,\alpha,\beta > 0$ and $\alpha+\beta < 1$. Again, nothing is assumed about the true orders (p_A, q_A) and only the standard assumptions are posed for the true link function h_A.

Theorem 5 *Let* $\{Y_t, \lambda_{t,A}\}$ *and* $\{\boldsymbol{X}_t\}$ *satisfy* $\mathcal{A}2$ *with a link function* h_A*, true parameter* $\boldsymbol{\theta}_A$ *and distribution* $\mathcal{F}_\lambda$*. Let* h_0 *be given by* (19) *and* $\boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$ *be a compact set. Assume that* (17)*,* (18) *and* $\mathcal{A}7$ *hold. Then*

$$\frac{1}{T}\Delta_{T,W} \xrightarrow{\mathsf{P}} \int_0^1 \int_{\mathbb{R}^d} \zeta^2(u,\boldsymbol{v}) w(u)\omega(\boldsymbol{v})\mathrm{d}\boldsymbol{v}\mathrm{d}u$$

as $T \to \infty$, *where*

$$\zeta(u,\boldsymbol{v}) = \mathbb{E}\left\{\left[g_{\lambda_{1,A}}(u) - g_{\lambda_{1,0}(\boldsymbol{\theta}_0)}(u)\right]\alpha\big(\boldsymbol{Z}_{1,0}(\boldsymbol{\theta}_0),\boldsymbol{v}\big)\right\},$$

and $\{\lambda_{t,0}(\boldsymbol{\theta}_0))\}$ *is a stationary sequence such that*

$$\lambda_{t,0}(\boldsymbol{\theta}) = \frac{\omega}{1-\beta} + \sum_{j=0}^{\infty} \beta^j [\alpha Y_{t-1-j} + \boldsymbol{\gamma}^\top \boldsymbol{X}_{t-1-j}],$$

and $\boldsymbol{Z}_{t,0}(\boldsymbol{\theta}) = (Y_{t-1}, \lambda_{t-1,0}(\boldsymbol{\theta}), \boldsymbol{X}_{t-1}^\top)^\top$.

In order to further discuss the conditions of consistency, write $\Delta^{\zeta}_{\infty,W}$ for each probability limit of $\Delta_{T,W}/T$ figuring in Theorems 3, 4 and 5. Since ζ is a continuous function on $(0,1)\times\mathbb{R}^d$, $\Delta^{\zeta}_{\infty,W} = 0$ only if $\zeta(u,\boldsymbol{v})$ is equal to zero identically in $(u,\boldsymbol{v}) \in (0,1)\times\mathbb{R}^d$. On the other hand if $\zeta(u,\boldsymbol{v}) \neq 0$ for some $(u,\boldsymbol{v}) \in (0,1)\times\mathbb{R}^d$, then it follows from assumption $\mathcal{A}7$ that $\Delta^{\zeta}_{\infty,W} > 0$, and consequently the corresponding test that rejects the null hypothesis H_0 for large values of $\Delta_{T,W}$ is consistent, in the sense that the rejection probability is equal to one asymptotically under the violations of H_0 considered herein.

In this connection note that in Theorem 3 $g^A_\lambda \neq g_\lambda$, and therefore $g^A_{\lambda_1}(u) - g_{\lambda_1}(u)$ is generally a nonzero random variable. If in addition $g^A_\lambda(u) > g_\lambda(u)$ for all $u \in (0,1)$ (which is the case for $\mathcal{F}_\lambda$ Poisson and $\mathcal{F}^A_\lambda$ a Negative Binomial), then $g^A_{\lambda_1}(u) - g_{\lambda_1}(u) >$

0 for all $u \in (0,1)$. In the case of Theorems 4 and 5, if the condition (2) holds sharply in the sense that $\{\lambda_1 < \lambda_2 \implies g_{\lambda_1}(u) > g_{\lambda_2}(u)$ for all $u \in (0,1)\}$ (as, e.g., for the Poisson distribution), then $\lambda_{1,A} - \lambda_{1,0} \neq 0$ a.s. implies that $g_{\lambda_{1,A}}(u) - g_{\lambda_{1,0}}(u) \neq 0$ a.s. for all $u \in (0,1)$. Also observe that $\alpha(\boldsymbol{v}_1, \boldsymbol{v}_2) = 0$ only if $\boldsymbol{v}_1^\top \boldsymbol{v}_2 = (3\pi/4) + k\pi,\ k = 0, \pm 1, \pm 2, \ldots$. Hence in each of the three cases of alternatives considered by Theorems 3–5, the random function within the corresponding expectation in the definition of $\zeta(u, \boldsymbol{v})$ is generally nonzero. This of course does not rule out the case of $\zeta(\cdot,\cdot)$ being identically equal to zero, but alludes to the very special circumstances under which the test based on $\Delta_{T,W}$ will miss any of the alternatives considered herein, at least for large sample size T. In this connection, the small sample behavior of the test based on $\Delta_{T,W}$ under various realizations of the alternatives considered in Theorems 3–5 is investigated by means of a Monte Carlo simulation study in Sect. 6 and shows that the new test has non-trivial power against several instances of violations of the null hypothesis.

5 Practical computation and bootstrap test

5.1 Analytic computations

Let the weight function W be of the form $W(u, \boldsymbol{v}) = w(u)\omega(\boldsymbol{v})$. Due to (14) the test statistic $\Delta_{T,W}$ takes the form

$$\Delta_{T,W} = \frac{1}{T} \sum_{t,s=1}^{T} K_{w,ts} \widetilde{K}_{\omega,ts}, \tag{20}$$

with

$$K_{w,ts} = \int_0^1 \widehat{\varepsilon}_t(u)\widehat{\varepsilon}_s(u) w(u) \mathrm{d}u = \int_0^1 \left[u^{Y_t} - g_{\widehat{\lambda}_t}(u)\right]\left[u^{Y_s} - g_{\widehat{\lambda}_s}(u)\right] w(u) \mathrm{d}u,$$

and $\widetilde{K}_{\omega,ts} := \widetilde{K}_\omega(\widehat{\boldsymbol{Z}}_t - \widehat{\boldsymbol{Z}}_s)$, where $\widetilde{K}_\omega(\boldsymbol{z}) = \int_{\mathbb{R}^d} \cos(\boldsymbol{v}^\top \boldsymbol{z})\omega(\boldsymbol{v})\mathrm{d}\boldsymbol{v}$. The integral $\widetilde{K}_\omega(\cdot)$ can be easily analytically computed if the weight function $\omega(\cdot)$ corresponds to the density of a spherical distribution. Then we immediately have $\widetilde{K}_\omega(\boldsymbol{z}) = \Xi(\|\boldsymbol{z}\|)$ with the function $\Xi(\cdot)$ being the generator of the characteristic function of the chosen spherical distribution. Hereafter we use $\Xi := \Xi_{\gamma,\eta}$, with

$$\Xi_{\gamma,\eta}(\xi) = e^{-\gamma \xi^\eta},\ \xi, \gamma > 0,\ \eta \in (0,2],$$

corresponding to the spherical stable characteristic function. Notice that Assumption $\mathcal{A}7^*$ is satisfied for all $\gamma > 0$ and $\eta \in (0,2]$,, while only $\eta = 2$ fulfills the requirements of Assumption $\mathcal{A}7$.

The integral $K_{w,ts}$ can be further simplified depending on the PGF g_λ and the choice of $w(u)$. Commonly used function is $w_\rho(u) = u^\rho$ for $\rho \geq 0$. Another possibility is to take a weight function proportional to a density of a beta distribution $w_{\rho,\kappa}(u) =$

$u^{\rho}(1-u)^{\kappa}$ for $\rho, \kappa \geq 0$, which enables to regulate the weight given to neighborhoods of the origin and point $u = 1$. The case of linear Poisson PARX model and weight function w_{ρ} is considered in more detail in Sect. 5.3.

5.2 Bootstrap version of the test

Theorem 1 provides the asymptotic distribution of the test statistics $\Delta_{T,W}$ under the null hypothesis $\mathcal{H}_0$, but this distribution depends on several unknown quantities. To get an approximation for the desired critical value one can generate the Gaussian process described in Theorem 1 with the unknown quantities replaced by their estimators and then calculate the integral and its quantiles. However, this approach is typically computationally demanding, and appropriately chosen resampling methods have proved useful in similar setups.

In a model without external covariates, that is when $\pi \equiv 0$ in (5), a parametric bootstrap is a natural solution, see Neumann (2021). Indeed, under the null hypothesis the conditional distribution of Y_t given $\mathcal{I}_{t-1}$ is completely specified up to the finite dimensional parameter $\boldsymbol{\theta}$. Given an estimator $\widehat{\boldsymbol{\theta}}$, one can recursively generate bootstrap copies of the series and compute bootstrap replications of the test statistic $\Delta_{T,W}^{(b)}$ for $b = 1, \ldots, B$. Analogous approach can be used if $\boldsymbol{X}_t$ consists solely from deterministic periodic components in a sense of Remark 5. Then one can successively generate bootstrap series $\{Y_t^{(*b)}\}_{t=1}^{T}$, for fixed $\{\boldsymbol{X}_t\}_{t=0}^{T-1}$, and the parametric bootstrap is analogous to the parametric residual bootstrap in a regression. If $\boldsymbol{X}_t$ is fully stochastic, then the same approach (taking fixed $\{\boldsymbol{X}_t\}_{t=0}^{T-1}$) can be applied as well, leading to an inference which could be understood as conditional on this observed covariate process. A slightly more general approach, which takes into account the stochastic nature of $\{\boldsymbol{X}_t\}_{t=0}^{T-1}$, performs also resampling of this covariate process, leading to bootstrap versions $\{\boldsymbol{X}_t^{(*b)}\}_{t=0}^{T-1}$ used for generating the bootstrap outcome series $\{Y_t^{(*b)}\}_{t=1}^{T}$. Since the observations in $\{\boldsymbol{X}_t\}$ are generally correlated, a block bootstrap by Künch (1989) is suitable here, as it mimics the original structure of the data while generating resamples. It proceeds as follows: The original series $\{\boldsymbol{X}_t\}_{t=0}^{T-1}$ is divided into blocks of length l of successive observations and then a random sampling with replacement is used for these blocks. The blocks can be taken either overlapping or non-overlapping. If the block length is $l = T$, then the method simplifies to the fixed regressors approach. Alternatively, one can use the stationary bootstrap of Politis and Romano (1994), where the resampling is applied to properly chosen blocks with a random length and which yields a stationary process $\{\boldsymbol{X}_t^{(*b)}\}_{t=0}^{T-1}$.

Hence, we propose a parametric bootstrap based on a certain combination of the bootstrap developed by Neumann (2021) adjusted for models with covariates by combining it with the stationary bootstrap by Politis and Romano (1994) or the classical block bootstrap by Künch (1989).

For given the observed data $(Y_1, \ldots, Y_T; \boldsymbol{X}_0, \ldots, \boldsymbol{X}_{T-1})$ and specified weight function W the test proceeds as follows:

1. Calculate the estimator $\widehat{\boldsymbol{\theta}}_T$ of $\boldsymbol{\theta}$ from data $(Y_1, \ldots, Y_T; \boldsymbol{X}_0, \ldots, \boldsymbol{X}_{T-1})$ and the corresponding value of the test statistic $\Delta_{T,W}$.

2. For $b = 1, \dots, B$ do:
   (i) Generate bootstrap observations $\boldsymbol{X}_0^{(*b)}, \dots, \boldsymbol{X}_{T-1}^{(*b)}$ from $\boldsymbol{X}_0, \dots, \boldsymbol{X}_{T-1}$.
   (ii) Generate the bootstrap observations $Y_1^{(*b)}, \dots, Y_T^{(*b)}$ recursively such that $Y_t^{(*b)}$ is generated from the distribution $\mathcal{F}_{\widehat{\lambda}_t^{(*b)}}$, where

$$\widehat{\lambda}_t^{(*b)} = r(Y_{t-1}^{(*b)}, \dots, Y_{t-p}^{(*b)}; \widehat{\boldsymbol{\theta}}_1) + \pi(\boldsymbol{X}_{t-1}^{(*b)}; \widehat{\boldsymbol{\theta}}_2).$$

   (iii) Calculate the estimator $\widehat{\boldsymbol{\theta}}_T^{(*b)}$ of $\boldsymbol{\theta}$ and the bootstrap version of the test statistic $\Delta_{T,W}^{(*b)}$ from data $(Y_1^{(*b)}, \dots, Y_T^{(*b)}; \boldsymbol{X}_0^{(*b)}, \dots, \boldsymbol{X}_{T-1}^{(*b)})$.
3. Compute the p-value of the test as

$$p = \frac{1}{B+1} \sum_{b=1}^{B} \mathrm{I}[\Delta_{T,W}^{(*b)} \geq \Delta_{T,W}].$$

The steps 1. and (iii) require initial values for computation of $\widehat{\lambda}_t$ and $\widehat{\lambda}_t^{(*b)}$. One possibility, used also in Sect. 6, is to take $\widehat{\lambda}_t = 0$, $Y_t = Y_1$ for $t \leq 0$.

As mentioned above, in (ii) we can take either $\boldsymbol{X}_t^{(*b)} = \boldsymbol{X}_t$ (no bootstrapping of covariates), or apply one of the suitable resampling methods, of combination of both in a case when part of the covariates are deterministic and the rest is stochastic, see Section 7 for an example. The simulation study in the next section uses overlapping block bootstrap with fixed block length l with $l \approx n^{1/3}$.

Using similar tools as in Neumann (2021) while examining step–by–step the proof of Theorem 1, it is expected that the approximation of the critical value by the above parametric bootstrap is asymptotically correct if the original data follow the null hypothesis. This claim is supported by the results of the Monte Carlo simulations conducted in Sect. 6. However, a detail proof of this statement is beyond the scope of this paper.

5.3 Application to PARX model

The linear Poisson autoregression with exogenous covariates (PARX) assumes that $\mathcal{F}_\lambda$ is Poisson, so the corresponding PGF takes form $g_\lambda(u) = e^{\lambda(u-1)}$, and the conditional mean is modelled as (9) for some $p \geq 1$ and $q \geq 0$ and an unspecified model parameter $\boldsymbol{\theta}$. In the following this particular hypothesis is referred to as $\mathcal{H}_0^{\mathrm{P}}$. Then $\widehat{\varepsilon}_t(u) = u^{Y_t} - e^{\widehat{\lambda}_t(u-1)}$, where $\widehat{\lambda}_t$ are computed recursively as described in Sect. 3. The integral $K_{w,ts}$ from (20) can be computed as $K_{w,ts} := K_w(\widehat{\varepsilon}_t, \widehat{\varepsilon}_s)$ for

$$K_w(\varepsilon_1, \varepsilon_2) = \int_0^1 (u^{y_1} - e^{\lambda_1(u-1)})(u^{y_2} - e^{\lambda_2(u-1)}) w(u) \mathrm{d}u,$$

where λ_t and y_t, $t = 1, 2$, are used as generic notations for pairs of parameters and associated responses, respectively. To proceed further hereafter we adopt the weight function $w(u) = w_\rho(u) = u^\rho$, $\rho \geq 0$, and write $K^{(\rho)}(\varepsilon_1, \varepsilon_2) := K^{(\rho)}(y_1, y_2; \lambda_1, \lambda_2)$, for the resulting integral which after some straightforward algebra is rendered as

$$\begin{aligned} K^{(\rho)}(y_1, y_2; \lambda_1, \lambda_2) = &\frac{1}{1 + y_1 + y_2 + \rho} \\ &- e^{-\lambda_2} I(y_1 + \rho, \lambda_2) - e^{-\lambda_1} I(y_2 + \rho, \lambda_1) \\ &+ e^{-(\lambda_1+\lambda_2)} I(\rho, \lambda_1 + \lambda_2), \end{aligned}$$

where $I(a, \theta) = \int_0^1 u^a e^{\theta u} \mathrm{d}u = \frac{(-\theta)^{-a}}{\theta} \left(\Gamma(1 + a, -\theta) - \Gamma(1 + a)\right)$. Then the test statistic figuring in (20) may be written as

$$\Delta_{T,W} = \frac{1}{T} \sum_{t,s=1}^{T} \widehat{K}_{ts}^{(\rho)} \Xi_{\gamma,\eta}(\|\widehat{\boldsymbol{Z}}_t - \widehat{\boldsymbol{Z}}_s\|),$$

with $\widehat{K}_{ts}^{(\rho)} = K^{(\rho)}(y_t, y_s; \widehat{\lambda}_t, \widehat{\lambda}_s)$ and $\widehat{\boldsymbol{Z}}_t$ defined recursively as described in Sect. 3.

The weight function w in general allows one to assign different weights to different points $u \in [0, 1]$ in the definition of $\Delta_{T,W}$. In the context of PARX models without covariates, Hudecová et al. (2015) considered a different weighted L_2-type GOF test statistic and employed the same weighting function $w_\rho(u) = u^\rho$. Their Monte Carlo simulations show that test statistics corresponding to different choices of ρ yield different powers against different alternatives, and it is not possible to choose a fixed ρ which would lead to the largest power against all kinds of alternatives. Exploration of other weighting functions and the problem of an optimal choice of ρ remains an open problem. Clearly, the simplest case is $\rho = 0$, i.e., no weighting. In the simulations in the next section we take this simplest choice for w and focus rather on a comparison of powers of $\Delta_{T,W}$ for different ω functions.

6 Simulations

This simulation study explores the finite sample behavior of the suggested bootstrap test based on $\Delta_{T,W}$ for the null hypothesis $\mathcal{H}_0^{\mathrm{P}}$, where the conditional distribution $\mathcal{F}_\lambda$ is Poisson. The conditional mean is specified either as an ARX(1) kind of model

$$\lambda_t = \omega + \alpha_1 Y_{t-1} + c\,(\cos X_{t-1} + 1)\,, \tag{S1}$$

or as a GARCHX(1,1) type of model

$$\lambda_t = \omega + \alpha_1 Y_{t-1} + \beta_1 \lambda_{t-1} + c\,(\cos X_{t-1} + 1)\,. \tag{S2}$$

The exogenous series $\{X_t\}$ was generated from a stationary AR(1) model $X_t = \rho_X X_{t-1} + \varepsilon_t$ for $\rho_X = 0.5$ and $\{\varepsilon_t\}$ iid centered normal with variance $(1 - \rho_X^2)^{-1}$. The model parameters were estimated by the conditional maximum likelihood method as proposed by Agosto et al. (2016), so the estimator is the same as obtained by the

Poisson QML from Aknouche and Francq (2021). The performance of the testing procedure was investigated under the null hypothesis. Moreover the alternative hypothesis settings considered are the cases whereby we have:

(1) Incorrect model specification. Observations were generated from a model of an incorrect order, while the function π from (5) and the conditional distribution $\mathcal{F}_\lambda$ were specified correctly. Namely, the test for (S1) was conducted for observations from the GARCHX(1,1) model (S2) and from a ARX(2) kind of model

$$\lambda_t = \omega + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + c\,(\cos X_{t-1} + 1)\,, \tag{S3}$$

which was used as an alternative for testing (S2) as well.

(2) Incorrect conditional distribution. Observations were generated from a model with a correctly specified conditional mean λ_t, but the conditional distribution was the Negative Binomial $\mathsf{NB}(\lambda_t, r)$ with a dispersion parameter r. Recall that if a random variable follows $\mathsf{NB}(\lambda, r)$ then its mean is λ and variance is $\lambda(1 + \lambda/r)$.

(3) Incorrectly specified dependence on the exogenous variable X_t. Observations were generated from a Poisson model with a correct orders p, q and $\pi(X_{t-1}; c) = c(\sin X_{t-1} + 1)$ (abbreviated as "sin" alternative) or $\pi(X_{t-1}; c) = c(2 - |X_{t-1}|)I[|X_{t-1}| < 2]$ (abbreviated as "abs" alternative) and we test for (S1) or (S2), where $\pi(X_{t-1}; c) = c(\cos X_{t-1} + 1)$.

Results are presented for sample sizes $T = 100, 200, 500$ and $T = 1\,000$ at level of significance $\alpha = 0.05$. The test statistic $\Delta_{T,W}$ is computed for a weight function $W(u, \boldsymbol{v}) = w(u)\omega(\boldsymbol{v})$, where $\omega(\cdot) = \omega_{\gamma,\eta}(\cdot)$ depends on tuning parameters $\gamma > 0$, $\eta \in (0, 2]$ as described in Sect. 5.3. As mentioned at the end of the previous section, we focus on exploration of the impact of ω on the power, so we take $w(u) = 1$, so there is simply no weighting with respect to u, while the parameters γ, η were considered in the following combinations

$$(\gamma, \eta) \in \Big\{(1/4, 1/4)\,, (1/2, 1/2)\,, (1/2, 1)\,, (1, 1/2)\,, (1, 1), (1, 2), (2, 2)\Big\}, \tag{21}$$

leading to 7 different ω functions. For model (S1) all these settings satisfy the assumptions required by the weight function, but for (S2) we have theoretical justification only for the last two choices of (γ, η) due to the stronger requirements in Assumption $\mathcal{A}7$. Despite this fact, we explore the behavior of the bootstrap test for all the choices of (γ, η) listed above, so we can investigate how sensitive the procedure is to violation of Assumption $\mathcal{A}7$. In addition, the behavior of the test statistic $\Delta_{T,W}$ was compared to tests based on the test statistics $\Delta_T^{(0)} := \Delta_{T,w}^{(0)}$ for $w(u) = 1$ and $\Delta_T^{(1)}$ from Remark 4.

The significance of $\Delta_{T,W}$ and $\Delta_T^{(i)}$, $i = 0, 1$, was evaluated using the parametric bootstrap described in Sect. 5 with resampling size $B = 499$. The simulations were conducted in the R-computing environment for $M = 1\,000$ repetitions.

The size of the test in Table 1 is presented for the setting $\omega = 0.2$, $\alpha_1 = 0.3$, $\pi_1 = 0.5$ for (S1) and for $\omega = 0.2$, $\alpha_1 = 0.3$, $\beta_1 = 0.3$, $c = 0.5$ for (S2). The results show that the test based on $\Delta_{T,W}$ and $\Delta_T^{(0)}$ keeps the prescribed confidence level.

Table 1 Size of the bootstrap test for λ_t specified by (S1) and (S2) for $\alpha = 0.05$

	$\mathcal{H}_0^{\mathrm{P}}$: ARX(1) model (S1)								
	$\Delta_{T,W}$ with $\rho = 0$ and (γ, η)								
T	$(\frac{1}{4}, \frac{1}{4})$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, 1)$	$(1, \frac{1}{2})$	(1,1)	(1,2)	(2,2)	$\Delta_T^{(0)}$	$\Delta_T^{(1)}$
100	0.054	0.050	0.040	0.041	0.046	0.045	0.046	0.052	0.027
200	0.036	0.043	0.053	0.054	0.057	0.052	0.061	0.038	0.037
500	0.047	0.050	0.049	0.048	0.055	0.054	0.057	0.048	0.025
1000	0.047	0.048	0.048	0.051	0.053	0.049	0.043	0.050	0.046
	$\mathcal{H}_0^{\mathrm{P}}$: GARCHX(1,1) model (S2)								
	$\Delta_{T,W}$ with $\rho = 0$ and (γ, η)								
T	$(\frac{1}{4}, \frac{1}{4})$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, 1)$	$(1, \frac{1}{2})$	(1,1)	(1,2)	(2,2)	$\Delta_T^{(0)}$	$\Delta_T^{(1)}$
100	0.040	0.041	0.037	0.042	0.037	0.039	0.039	0.041	0.032
200	0.051	0.043	0.038	0.041	0.035	0.028	0.030	0.053	0.046
500	0.049	0.050	0.048	0.046	0.039	0.049	0.049	0.054	0.055
1000	0.048	0.041	0.041	0.039	0.040	0.040	0.041	0.041	0.055

The test corresponding to $\Delta_T^{(1)}$ seems to be slightly conservative for (S1) model with $T < 1\,000$.

The power of the test for (S1) in Table 2 was computed for data generated from a ARX(2) type of model (S3) with $\omega = 0.2$, $\alpha_1 = 0.3$, $\alpha_2 = 0.6$, $c = 0.5$, a GARCHX(1,1) model (S2) with $\omega = 0.2$, $\alpha_1 = 0.3$, $\beta_1 = 0.6$, $c = 0.5$, a Negative Binomial model with $r = 3$ and λ_t given by (S1) for $\omega = 0.2$, $\alpha_1 = 0.3$, $c = 0.5$, and for a model with misspecified π with $\omega = 0.2$, $\alpha_1 = 0.3$, $c = 0.5$.

A first observation is that the test based on $\Delta_{T,W}$ clearly outperforms its competitors for the "sin" misspecification alternative of the π function regardless of the choice of the tuning parameters γ and η. Otherwise the new test appears to be often sensitive to the choice of γ and η, with the choices $(\gamma, \eta) = (1/4, 1/4)$, and $(\gamma, \eta) = (1/2, 1/2)$ being overall preferable. In this connection, the test based on $(\gamma, \eta) = (1/4, 1/4)$, outperforms the tests based on $\Delta_T^{(0)}$, and $\Delta_T^{(1)}$, nearly uniformly over the sample size T and against all alternatives, but often the power differential is small. The test corresponding to $(\gamma, \eta) = (1/2, 1/2)$ also performs well against competitors. In fact in the case of the "sin" alternative, even larger values of the pair (γ, η) lead to improved power, but at the same time all tests find it rather hard to distinguish an "abs" misspecification alternative of the π function.

The power of test for (S2) for the considered alternatives is shown in Table 3. The data were generated from an ARX(2) model (S3) with $\omega = 0.2$, $\alpha_1 = 0.3$, $\alpha_2 = 0.6$, and $c = 0.5$, from a Negative Binomial GARCHX(1,1) model with $r = 3$ and a pair of GARCHX(1,1) models with misspecified ψ function ("sin" and "abs") for $\omega = 0.2$, $\alpha_1 = 0.3$, $\beta_1 = 0.3$, $c = 0.5$.

The results in Table 3 are quantitatively similar with those of Table 2 for the NB distributional alternative and the "abs" misspecification of the π function, and qualitatively similar for the other two types of alternatives, but in this case it is harder to identify a ARX(2) alternative, while the power against a "sin" misspecification of the π function is somewhat lower for all tests. Comparisonwise, the conclusions are

Table 2 Power of test for ARX(1) model (S1) for data generated from ARX(2) model (S3), data from GARCHX(1,1) model (S2), data from the Negative Binomial distribution with $r = 3$ (NB), and data from incorrectly specified link part π, where the data are generated using $\pi(x; c) = c(\sin x + 1)$ (sin) or $\pi(x; c) = c(2 - |x|)I[|x| < 2]$ (abs), while we test $\pi(x; c) = c(\cos x + 1)$

		$\Delta_{T,W}$ with $\rho = 0$ and (γ, η)								
Alt.	T	$(\frac{1}{4}, \frac{1}{4})$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, 1)$	$(1, \frac{1}{2})$	(1,1)	(1,2)	(2,2)	$\Delta_T^{(0)}$	$\Delta_T^{(1)}$
ARX(2)	100	0.485	0.462	0.439	0.397	0.379	0.346	0.318	0.365	0.446
	200	0.772	0.729	0.684	0.643	0.595	0.554	0.518	0.687	0.767
	500	0.998	0.989	0.969	0.967	0.934	0.886	0.859	0.985	1.000
	1000	1.000	1.000	0.999	1.000	0.998	0.991	0.990	1.000	1.000
GARCHX	100	0.102	0.097	0.088	0.081	0.070	0.065	0.062	0.045	0.061
	200	0.154	0.158	0.146	0.128	0.116	0.113	0.104	0.116	0.137
	500	0.310	0.309	0.259	0.251	0.219	0.182	0.171	0.326	0.317
	1000	0.576	0.539	0.386	0.440	0.314	0.242	0.226	0.534	0.600
NB	100	0.639	0.632	0.575	0.609	0.460	0.318	0.263	0.618	0.522
	200	0.889	0.884	0.841	0.858	0.705	0.532	0.408	0.881	0.812
	500	1.000	1.000	0.999	0.999	0.987	0.911	0.795	1.000	0.995
	1000	1.000	1.000	1.000	1.000	1.000	0.999	0.989	1.000	1.000
sin	100	0.153	0.534	0.728	0.675	0.738	0.724	0.594	0.099	0.089
	200	0.343	0.893	0.972	0.953	0.969	0.968	0.914	0.163	0.145
	500	0.887	1.000	1.000	1.000	1.000	1.000	1.000	0.292	0.283
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.608	0.546
abs	100	0.058	0.054	0.053	0.057	0.064	0.059	0.061	0.049	0.047
	200	0.058	0.062	0.077	0.084	0.091	0.101	0.103	0.054	0.054
	500	0.076	0.091	0.091	0.110	0.126	0.116	0.130	0.071	0.065
	1000	0.095	0.110	0.128	0.162	0.190	0.232	0.235	0.103	0.087

also analogous to those drawn from the results of Table 2, overall favoring the tests based on $\Delta_{T,W}$, with $\gamma = \eta = 1/4$ and $\gamma = \eta = 1/2$. It is interesting to see that the procedure leads to reasonable results even if $\eta < 2$. The corresponding bootstrap test keeps the prescribed level, and this choice seems to be even preferable against some alternatives.

7 Real data analysis

Road crashes have not only a devastating impact on victims and their families, but they also bear a heavy economic cost, see, e.g., (Peden et al. 2004, Table 2.9). The total cost of motor vehicle collisions involve lost productivity, medical costs, legal and court costs, emergency services, insurance administration, travel delay, property damage, and workplace losses (Blincoe et al. 2015). Among all accidents, injuries to pedestrians and bicyclists caused 7 % of the total economic costs and 10 % of the total societal harm in USA in 2010 (Blincoe et al. 2015).

It is well known that weather affects both the driver capabilities and road conditions via visibility impairments, precipitation, and temperature extremes. In this application, we model daily numbers of accidents involving at least one pedestrian in Prague, the capital of the Czech Republic, in years 2018–2019. The aim is to built a time series

Table 3 Power of test for GARCHX(1,1) model (S2) for data generated from ARX(2) model (S3), data from the Negative Binomial distribution with $r = 3$ (NB), and data from incorrectly specified π, where the data are generated using $\pi(x; c) = c(\sin x + 1)$ (sin) or $\pi(x; c) = c(2 - |x|)I[|x| < 2]$ (abs), while we test $\pi(x; c) = c(\cos x + 1)$

		$\Delta_{T,W}$ with $\rho = 0$ and (γ, η)								
Alt.	T	$(\frac{1}{4}, \frac{1}{4})$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, 1)$	$(1, \frac{1}{2})$	(1,1)	(1,2)	(2,2)	$\Delta_T^{(0)}$	$\Delta_T^{(1)}$
ARX(2)	100	0.057	0.046	0.046	0.031	0.023	0.017	0.018	0.039	0.023
	200	0.115	0.095	0.074	0.056	0.038	0.025	0.018	0.093	0.063
	500	0.291	0.274	0.229	0.171	0.128	0.063	0.054	0.268	0.243
	1000	0.545	0.602	0.571	0.484	0.270	0.212	0.148	0.508	0.243
NB	100	0.893	0.898	0.872	0.894	0.838	0.723	0.696	0.873	0.795
	200	0.997	0.996	0.994	0.996	0.988	0.923	0.905	0.996	0.978
	500	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
sin	100	0.075	0.172	0.400	0.321	0.421	0.375	0.258	0.062	0.056
	200	0.140	0.444	0.730	0.663	0.750	0.696	0.554	0.098	0.093
	500	0.355	0.935	0.995	0.991	1.000	0.991	0.953	0.213	0.191
	1000	0.741	1.000	1.000	1.000	1.000	1.000	1.000	0.388	0.191
abs	100	0.055	0.062	0.056	0.058	0.054	0.052	0.044	0.054	0.050
	200	0.050	0.053	0.051	0.047	0.049	0.055	0.043	0.051	0.057
	500	0.064	0.065	0.063	0.068	0.066	0.076	0.075	0.054	0.055
	1000	0.065	0.066	0.071	0.073	0.082	0.099	0.091	0.059	0.055

model for the number of collisions that allows dependence on variables reflecting the actual weather, namely the average daily temperature (AT) and precipitation (P). Beside the weather variables, also the weekly and yearly patterns need to be taken into account. Details regarding the data set are provided in the Supplementary file.

Various transformations of the explanatory variables were considered. The conditional distribution $\mathcal{F}_\lambda$ was assumed to be Poisson. Competitive models were fitted by Poisson ML and compared using information criteria, similarly as in Agosto et al. (2016). The following model was chosen as the best one:

$$\lambda_t = \underset{(0.101)}{0.572} + \underset{(0.034)}{0.020}Y_{t-1} + \underset{(0.101)}{0.967}\mathrm{W}_t + \underset{(0.066)}{0.318}\cos\left(\frac{2\pi t}{365}\right) + \underset{(0.097)}{0.229}\mathrm{I}(\mathrm{P}_t > 0), \tag{22}$$

where the numbers in parentheses are the standard deviations. Note that if $\mathcal{F}_\lambda$ was not Poisson, then the estimates in (22) are the Poisson QMLE, which are still valid under mild regularity conditions, see Aknouche and Francq (2021), (however, with different standard deviations).

Since the assumption of the Poisson distribution can be questionable in applications, we conduct the goodness-of-fit test. It is visible that the coefficient next to Y_{t-1} is not statistically significant, which indicates that the explanatory variables might explain all temporary dependence in the series. However, for illustrative purposes, we keep this term in the model and conduct the goodness-of-fit test for the null hypothesis $\mathcal{H}_0^P : Y_t|\mathcal{I}_{t-1} \sim \mathsf{Po}(\lambda_t)$, with $\lambda_t = \omega + \alpha Y_{t-1} + \gamma_1 \mathrm{W}_t + \gamma_2 \cos\left(\frac{2\pi t}{365}\right) + \gamma_3 \mathrm{I}(\mathrm{P}_t > 0)$ for some parameter $(\omega, \alpha, \gamma_1, \gamma_2, \gamma_3)^\top$. The test statistic $\Delta_{T,W}$ and its significance were computed as described in Sect. 5. Namely, we made use of the same weighting function ω as in Sect. 6 for the same set of tuning parameters (γ, η) as in (21). The bootstrap goodness-of-fit model was performed for $B = 1000$, where the block bootstrap was applied solely to the stochastic variable $\mathrm{I}(\mathrm{P}_t > 0)$, with block length $l = 10$, while the deterministic variables are held fixed. The obtained p-values for (γ, η) as in (21) are 0.132, 0.114, 0.154, 0.115, 0.175, 0.220, 0.219, respectively. Since none of the obtained p-values is significant, the Poisson distributional assumption seems to be plausible for these data. This is also supported by a probability integral transformation (PIT) diagnostic plot constructed as in Czado et al. (2009) and presented in the Supplementary file.

8 Discussion and conclusion

A goodness-of-fit test for count time series with exogenous covariates is proposed. The new test statistic is formulated in terms of Bieren's characterization and may be applied to general models of this type provided that the conditional PGF is specified. The limit distribution of this statistic is obtained under the null hypothesis of fixed, but otherwise arbitrary, order and rather general form of the link function. Asymptotic results for non-null situations are also presented, including consistency of the test against distributional misspecification as well as against certain types of violations of the hypothesized functional form of the conditional mean. The finite-sample performance of a bootstrap version of the suggested test is investigated by a Monte Carlo study against competing procedures, and corroborates the consistency features of our test. Finally the proposed methodology is used in order to perform explanatory analysis of the main factors affecting road accidents in the city of Prague.

Supplementary information

The online Supplementary file contains a description of the data set used in Sect. 7, some additional details about the analysis, implementation and proofs of all assertions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11749-024-00933-x.

Acknowledgements The research of Marie Hušková was partially supported by grant GAČR– 21–13323 S. The research of Šárka Hudecová was supported by grant GAČR 22–01639K.

Funding Open access funding provided by North-West University.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Agosto A, Cavaliere G, Kristensen D, Rahbek A (2016) Modeling corporate defaults: Poisson autoregressions with exogenous covariates. J Empir Finance 38:640–663

Agosto A, Giudici P (2020) A Poisson autoregressive model to understand COVID-19 contagion dynamics. Risks 8:77

Agosto A, Raffinetti E (2019) Validation of PARX models for default count prediction. Front Artif Intell 2:9

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

Aknouche A, Al-Eid E (2012) Asymptotic inference of unstable periodic ARCH processes. Stat Inference Stoch Process 15:61–79

Aknouche A, Francq C (2021) Count and duration time series with equal conditional stochastic and mean orders. Economet Theor 37:248–280

Angelini G, De Angelis L (2017) PARX model for football matches predictions. J Forecast 2:795–807

Baringhaus L, 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

Bentarzi M, Aries N (2022) QMLE of periodic integer-valued time series models. Comm Statist Simulation Comput 51(9):4973–4999

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

Bishop HR (1979) *The Use of the Empirical Probability Generating Function to Estimate the Neyman Type A Distribution Parameters*. Ph. D. thesis, Naval Postgraduate School, Monterey, CA

Blincoe LJ, Miller TR, Zaloshnja E, Lawrence BA (2015) The economic and societal impact of motor vehicle crashes, 2010. (revised). Technical Report DOT HS 812 013, National Highway Traffic Safety Administration, Washington, DC

Cavaliere G, Perera I, Rahbek A (2024) Specification tests for GARCH processes with nuisance parameters on the boundary. J Bus Econom Statist 42:197–214

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

Czado C, Gneiting T, Held L (2009) Predictive model assessment for count data. Biometrics 65:1254–1261

Escanciano JC, Velasco C (2006) Generalized spectral tests for the martingale difference hypothesis. J Econometrics 134:151–185

Esquível ML (2008) Probability generating functions for discrete real-valued random variables. Theory Probab Appl 52:40–57

Feuerverger A (1988) On the empirical saddlepoint approximation. Biometrika 76:457–464

Fokianos K, Neumann MH (2013) A goodness-of-fit tests for Poisson count processes. Electron J Stat 7:793–819

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

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

Francq C, Zakoïan J (2007) Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. Stochastic Process Appl 117:1265–1284

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

Hlávka Z, Hušková M, Meintanis SG (2017) Fourier-type tests involving martingale difference processes. Economet Rev 46:468–492

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

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

Hušková M, Meintanis SG (2010) Tests for the error distribution in nonparametric possibly heteroscedastic regression models. TEST 19:92–112

Jiménez-Gamero M, Batsidis A (2017) Minimum distance estimators for count data based on the probability generating function with applications. Metrika 80:503–545
Jiménez-Gamero MD (2014) On the empirical characteristic function process of the residuals in GARCH models and applications. TEST 23:409–432
Jiménez-Gamero MD, Lee S, Meintanis SG (2020) Goodness-of-fit tests for parametric specifications of conditionally heteroscedastic models. TEST 29:682–703
Johnson NL, Kemp AW, Kotz S (2005) Univariate Discrete Distributions, 3rd edn. John Wiley & Sons Inc., Hoboken, New Jersey
Kedem B, Fokianos K (2002) Regression Models for Time Series Analysis. John Wiley & Sons Inc, Hoboken, New Jersey
Künch H (1989) The jackknife and the bootstrap for general stationary observations. Ann Stat 17(428):1217–1241
Liboschik T, Fokianos K, Fried R (2017) tscount: An R package for analysis of count time series following generalized linear models. J Stat Softw 82:1–51
Meintanis SG, Bassiakos Y (2005) Goodness-of-fitt tests for additively closed count models with an application to the generalized Hermite distribution. Sankhya 67:538–552
Meintanis SG, Jiménez-Gamero MD, Alba-Fernandéz V (2014) A class of goodness-of-fit tests based on transformation. Comm Statist Theory Methods 43:1708–1735
Meintanis SG, Nikitin YY (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, Pérez-Abreu V (1993) Empirical probability generating function: an overview. Insurance Math Econom 12:287–295
Neumann MH (2021) Bootstrap for integer-valued GARCH(p, q) processes. Stat Neerl 75:343–363
Novoa-Muñoz F, Jiménez-Gamero MD (2014) Testing for the bivariate Poisson distribution. Metrika 77:771–793
Novoa-Muñoz F, Jiménez-Gamero MD (2016) A goodness-of-fit test for the multivariate Poisson distribution. SORT 40:113–138
Peden M, Scurfield R, Sleet D, Mohan D, Hyder AA, Jarawan E, Mathers C (2004) World report on road traffic injury prevention. Technical report, World Health Organization, Geneva
Politis DN, Romano JP (1994) The stationary bootstrap. J Amer Statist Assoc 89(428):1303–1313
Prezotti Filho PR, Reisen VA, Bondon P, Ispány M, Melo MM, Serpa FS (2021) A periodic and seasonal statistical model for non-negative integer-valued time series with an application to dispensed medications in respiratory diseases. Appl Math Model 96:545–558
Rémillard B, Theodorescu R (2000) Inference based on the empirical probability generating function for mixtures of Poisson distributions. Stat Risk Model 18:349–366
Rueda R, Pérez-Abreu V, O'Reilly F (1991) Goodness of fit for the Poisson distribution based on the probability generating function. Comm Statist Theory Methods 20:3093–3110
Schweer S (2016) A goodness-of-fit test for integer-valued autoregressive processes. J Time Series Anal 37:77–98
Szűcs G (2005) Approximations of empirical probability generating processes. Stat Risk Model 23:67–80
Weiss CH, Schweer S (2015) Detecting overdispersion in INARCH(1) processes. Stat Neerl 69:281–297
White H (1982) Maximum likelihood estimation of misspecified models. Econometrica 50:3093–3110
Zayed AI (1996) Handbook of Function and Generalized Function Transformations. CRC Press, New York
Zhu F (2011) A negative binomial integer-valued GARCH model. J Time Series Anal 32:54–67