Received: 23 July 2020 | Revised: 23 December 2020 | Accepted: 12 February 2021

DOI: 10.1111/stan.12238

ORIGINAL ARTICLE

WILEY

Bootstrap for integer-valued GARCH(p, q) processes

Michael H. Neumann

Friedrich-Schiller-Universität, Institut für Mathematik, Jena, Germany

Correspondence
Michael H. Neumann,
Friedrich-Schiller-Universität, Institut für Mathematik, 07743 Jena, Germany.
Email: michael.neumann@uni-jena.de

Abstract

We consider integer-valued processes with a linear or nonlinear generalized autoregressive conditional heteroscedastic models structure, where the count variables given the past follow a Poisson distribution. We show that a contraction condition imposed on the intensity function yields a contraction property of the Markov kernel of the process. This allows almost effortless proofs of the existence and uniqueness of a stationary distribution as well as of absolute regularity of the count process. As our main result, we construct a coupling of the original process and a model-based bootstrap counterpart. Using a contraction property of the Markov kernel of the coupled process we obtain bootstrap consistency for different types of statistics.

KEYWORDS

absolute regularity, bootstrap, coupling, integer-valued GARCH, mixing, stationarity

MOS SUBJECT CLASSIFICATION

Primary 60G10; secondary 60J05

[Correction added on 26 April 2021 after first online publication: Article title has been corrected in this version.]

1 | INTRODUCTION

Conditionally heteroscedastic processes have become quite popular for modeling the evolution of stock prices, exchange rates and interest rates. Starting with the seminal papers by Engle (1982) on autoregressive conditional heteroscedastic models (ARCH) and Bollerslev (1986) on generalized ARCH (GARCH), numerous variants of these models have been proposed for modeling financial time series; see for example, Francq and Zakoι¨an (2010) for a detailed overview. More recently, integer-valued GARCH models (INGARCH) which mirror the structure of GARCH models have been proposed for modeling time series of counts; see for example, Fokianos (2012) and the recently edited volume by Davis, Holan, Lund, and Ravishanker (2016).

We consider integer-valued GARCH processes of order p and q (INGARCH(p, q)). Processes of this type have enjoyed increasing popularity in the recent years. In most cases, it is assumed that the count process is observable and that, given the past, the count variables have a Poisson distribution with a random intensity which depends on lagged values of the count and the intensity process. Mixing properties of such processes have been derived for a first time in Neumann (2011), for INGARCH(1,1) processes under a contraction condition on the intensity function. This has been generalized to INGARCH processes of higher order in Aknouche and Francq (2018) as well as in Doukhan and Neumann (2019), in the latter paper under a weaker semi-contractive condition leading to a subexponential decay of the mixing coefficients only. Doukhan, Leucht, and Neumann (2020) proved absolute regularity of the count process again in the INGARCH(1,1) case but allowing a possibly nonstationary (explosive) bahavior of the process. Model-based bootstrap for INGARCH(1,1) processes has been used in Fokianos and Neumann (2013) and in Leucht and Neumann (2013, Section 5.3) in order to determine an appropriate critical value for goodness-of-fit tests. In the latter paper, the authors showed in particular that there exists a coupling of the original count process and its bootstrap counterpart which leads to bootstrap consistency for statistics of Cramér–von Mises type; see their Lemma 4. In the present contribution, we generalize these results in several directions. We consider the more general case of INGARCH processes of arbitrary finite order. More importantly, we derive general properties of the bootstrap process which can be used to derive consistency for statistics of different types.

In Section 2 we derive basic properties of the bivariate process consisting of the count process $(X_t)_t$ and the accompanying process $(\lambda_t)_t$ of random intensities. Note that in our context $(Z_t)_t$ with $Z_t = (X_t, \ldots, X_{t-p+1}, \lambda_t, \ldots, \lambda_{t-q+1})$ is a time-homogeneous Markov chain. We first show that a contraction condition for the intensity process leads to a contraction property for the conditional distribution of Z_t given Z_{t-1}. Being then in a Markovian context we show that this yields by the Banach fixed point theorem existence and uniqueness of a stationary distribution. Moreover, we exploit the contraction property once more to show absolute regularity of the count process. In Section 3 we introduce our model-based bootstrap method. In contrast to the majority of the existing papers on bootstrap consistency, we do not only justify this method for some particular statistic of interest. To achieve greater generality, we first show that the bootstrap process $(X_t^*)_{t=1,\ldots,n}$ mimics the stochastic behavior of the original count process. This is accomplished by a coupling of both processes on a suitable new probability space. We show, for these coupled processes, proximity of the original and the bootstrap version as well as absolute regularity. In Section 4 we demonstrate by some examples that the coupling result can be used to derive almost effortlessly the consistency of the bootstrap approximation for statistics which are of real interest. All proofs are deferred to a final Section 5.

2 | INGARCH(P, Q) PROCESSES AND THEIR PROPERTIES

2.1 | Notation and assumptions

We consider a class of integer-valued processes $(X_t)_{t\in\mathbb{Z}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where, for $t \in \mathbb{Z}$,

$$X_t|\mathcal{F}_{t-1} \sim \operatorname{Pois}(\lambda_t), \tag{1a}$$

$$\lambda_t = f_\theta(X_{t-1}, \dots, X_{t-p}, \lambda_{t-1}, \dots, \lambda_{t-q}), \tag{1b}$$

and $\mathcal{F}_s = \sigma(X_s, \lambda_s, X_{s-1}, \lambda_{s-1}, \dots)$ denotes the σ-field generated by the random variables up to time s. The parameter θ lies in Θ, the set of possible parameters, and $\theta_0 \in \Theta$ denotes the true parameter. A frequently considered special case is that of a linear model of order p, q, where

$$f_\theta(x_1, \dots, x_p, \lambda_1, \dots, \lambda_q) = \omega + \sum_{i=1}^{p} \alpha_i x_i + \sum_{j=1}^{q} \beta_j \lambda_j,$$

with $\theta = (\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$.

The process $(Z_t)_{t\in\mathbb{Z}}$ with $Z_t = (X_t, \dots, X_{t-p+1}, \lambda_t, \dots, \lambda_{t-q+1})$ is a time-homogeneous Markov chain with state space $S = \mathbb{N}_0^p \times [0, \infty)^q$. The following condition ensures existence and uniqueness of a stationary distribution of $(Z_t)_{t\in\mathbb{Z}}$ as well as absolute regularity of the count process $(X_t)_{t\in\mathbb{Z}}$. With a view towards a model-based bootstrap, where an estimator of θ is used instead of the true value, we require uniformity over some subset Θ_0 of Θ.

(A1) There exist nonnegative constants $c_1, \dots, c_p, d_1, \dots, d_q$ with $L := \sum_{i=1}^p c_i + \sum_{j=1}^q d_j < 1$ and some $\delta > 0$ such that

$$\left|f_\theta(x_1, \dots, x_p, \lambda_1, \dots, \lambda_q) - f_\theta(x_1', \dots, x_p', \lambda_1', \dots, \lambda_q')\right| \le \sum_{i=1}^{p} c_i\ |x_i - x_i'| + \sum_{j=1}^{q} d_j\ |\lambda_j - \lambda_j'|$$

holds for all $(x_1, \dots, x_p, \lambda_1, \dots, \lambda_q), (x_1', \dots, x_p', \lambda_1', \dots, \lambda_q') \in S$ and all $\theta \in \Theta_0$. Furthermore, we suppose that $C^{(0)} := \sup_{\theta\in\Theta_0} \{f_\theta(0, \dots, 0)\} < \infty$.

2.2 | Stationarity and finiteness of moments

For contractive INGARCH(1,1) processes, existence and uniqueness of a stationary distribution and absolute regularity of the count process has been first proved in Neumann (2011). Aknouche and Francq (2018) generalized these results to contractive INGARCH processes of higher order. Their proof of the existence and uniqueness of a stationary distribution is based on approximation techniques. In the present paper, our approach is a different one. We show in Proposition 1 below that (A1) yields a contraction property of the Markov kernel connected with $(Z_t)_t$ in terms of a suitable Wasserstein metric. As described in Eberle (2020, Chapter 3), existence and uniqueness of a stationary distribution follows then by the Banach fixed point theorem. Furthermore, we exploit the contraction property to derive almost effortlessly absolute regularity of the count process in Section 2.3. As our main result, we construct in Section 3 a coupling of the original and the bootstrap process which shows the desired similarity of these processes. We exploit the contraction property once more to prove the existence and uniqueness of a stationary version of this joint process as well as absolute regularity of the joint count processes. To this end, we first transfer the contraction condition (A1) for the intensity process into a contraction property for the Markov kernel steering the process $(Z_t)_t$.

We consider the following metric on S:

$$\Delta\left((x_1, \ldots, x_p, \lambda_1, \ldots, \lambda_q), (x_1', \ldots, x_p', \lambda_1', \ldots, \lambda_q')\right) = \sum_{i=1}^{p} \gamma_i \, |x_i - x_i'| + \sum_{j=1}^{q} \delta_j \, |\lambda_j - \lambda_j'|,$$

where $\gamma_1, \ldots, \gamma_p, \delta_1, \ldots, \delta_q$ are strictly positive constants. Let $z = (x_1, \ldots, x_p, \lambda_1, \ldots, \lambda_q)$, $z' = (x_1', \ldots, x_p', \lambda_1', \ldots, \lambda_q') \in S$ be arbitrary. With an appropriate choice of $\gamma_1, \ldots, \gamma_p$, $\delta_1, \ldots, \delta_q$, we can construct random vectors $Z = (X, x_1, \ldots, x_{p-1}, \lambda, \lambda_1, \ldots, \lambda_{q-1})$ and $Z' = (X', x_1', \ldots, x_{p-1}', \lambda', \lambda_1', \ldots, \lambda_{q-1}')$ on a suitable probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ such that $Z \sim \mathbb{P}_\theta^{Z_t|Z_{t-1}=z}$, $Z' \sim \mathbb{P}_\theta^{Z_t|Z_{t-1}=z'}$, and $\tilde{E}\Delta(Z, Z') \leq \kappa \, \Delta(z, z')$ holds for all $\theta \in \Theta_0$ and some $\kappa < 1$. Actually, according to the model Equation (1b), we have to set $\lambda = f_\theta(x_1, \ldots, x_p, \lambda_1, \ldots, \lambda_q)$ and $\lambda' = f_\theta(x_1', \ldots, x_p', \lambda_1', \ldots, \lambda_q')$. Let $(N(u))_{u \geq 0}$ be a Poisson process with unit intensity. We set $X \sim N(\lambda)$ and $X' \sim N(\lambda')$, which implies that $Z \sim \mathbb{P}_\theta^{Z_t|Z_{t-1}=z}$ and $Z' \sim \mathbb{P}_\theta^{Z_t|Z_{t-1}=z'}$, as required. Furthermore, since $\tilde{E}|X - X'| = |\lambda - \lambda'|$ we obtain by (1b) that

$$\begin{aligned}
\tilde{E}\Delta(Z, Z') &= (\gamma_1 + \delta_1) \left| f_\theta(x_1, \ldots, x_p, \lambda_1, \ldots, \lambda_q) - f_\theta(x_1', \ldots, x_p', \lambda_1', \ldots, \lambda_q') \right| \\
&\quad + \sum_{i=2}^{p} \gamma_i \, |x_{i-1} - x_{i-1}'| + \sum_{j=2}^{q} \delta_j \, |\lambda_{j-1} - \lambda_{j-1}'| \\
&\leq (\gamma_1 + \delta_1) \left(\sum_{i=1}^{p} c_i \, |x_i - x_i'| + \sum_{j=1}^{q} d_j \, |\lambda_j - \lambda_j'| \right) + \sum_{i=2}^{p} \gamma_i \, |x_{i-1} - x_{i-1}'| + \sum_{j=2}^{q} \delta_j \, |\lambda_{j-1} - \lambda_{j-1}'|.
\end{aligned} \tag{2}$$

The desired relation of $\tilde{E}\Delta(Z, Z') \leq \kappa\Delta(z, z')$ would be guaranteed to hold if we find strictly positive $\gamma_1, \ldots, \gamma_p, \delta_1, \ldots, \delta_q$ such that the right-hand side of (2) is less than or equal to $\kappa\Delta(z, z') = \kappa \left(\sum_{i=1}^{p} \gamma_i |x_i - x_i'| + \sum_{j=1}^{q} \delta_j |\lambda_j - \lambda_j'| \right)$, for all $(x_1, \ldots, x_p, \lambda_1, \ldots, \lambda_q)$, $(x_1', \ldots, x_p', \lambda_1', \ldots, \lambda_q') \in S$. The following lemma builds a bridge from the contraction property (A1) for the intensity function to a contraction property for the conditional distribution of Z_t given Z_{t-1}.

Lemma 1. *Let $c_1, \ldots, c_p, d_1, \ldots, d_q$ be nonnegative constants such that $\sum_{i=1}^{p} c_i + \sum_{j=1}^{q} d_j < 1$. Then there exist strictly positive constants $\gamma_1, \ldots, \gamma_p, \delta_1, \ldots, \delta_q$ and some $\kappa < 1$ such that*

$$(\gamma_1 + \delta_1) \left(\sum_{i=1}^{p} c_i \, y_i + \sum_{j=1}^{q} d_j \, z_j \right) + \sum_{i=2}^{p} \gamma_i \, y_{i-1} + \sum_{j=2}^{q} \delta_j \, z_{j-1} \;\leq\; \kappa \left(\sum_{i=1}^{p} \gamma_i \, y_i + \sum_{j=1}^{q} \delta_j \, z_j \right), \tag{3}$$

holds for all $y_1, \ldots, y_p, z_1, \ldots, z_q \geq 0$.

Let π_θ^Z be the Markov kernel which transfers Z_{t-1} to Z_t and let $\tilde{\pi}_\theta^Z$ be the Markov kernel which provides the above coupling, that is, for the above random variables Z and Z' we have that $(Z, Z') \sim \tilde{\pi}_\theta^Z((z, z'), \cdot)$.

The following proposition provides the contraction property which will be instrumental for the proof of the existence and uniqueness of a stationary distribution as well as for the derivation of absolute regularity of the count process.

Proposition 1. *Suppose that condition (A1) is fulfilled and that $\theta \in \Theta_0$. Let $\gamma_1, \dots, \gamma_p, \delta_1, \dots, \delta_p$ and $\kappa < 1$ be chosen as in Lemma 1.*

(i) Let $z, z' \in S$ be arbitrary. If $(Z, Z') \sim \tilde{\pi}_\theta^Z((z, z'), \cdot)$, then

$$Z \sim \mathbb{P}_\theta^{Z_t | Z_{t-1} = z} \qquad \text{and} \qquad Z' \sim \mathbb{P}_\theta^{Z_t | Z_{t-1} = z'}$$

and

$$\tilde{E}\Delta(Z, Z') \leq \kappa \; \Delta(z, z').$$

(ii) Let $((\tilde{Z}_t, \tilde{Z}'_t))_{t\in\mathbb{Z}}$ be a Markov chain with transition kernel $\tilde{\pi}_\theta^Z$. Then

$$\tilde{E}\Delta\left(\tilde{Z}_t, \tilde{Z}'_t\right) \leq \kappa \; \tilde{E}\Delta\left(\tilde{Z}_{t-1}, \tilde{Z}'_{t-1}\right).$$

In order to derive stationarity properties of the process $(Z_t)_{t\in\mathbb{Z}}$, we further translate the contraction result in Proposition 1 into a contraction property of the corresponding distributions. For two probability measures Q and Q' on S, we define the Kantorovich distance based on the metric Δ (also known as Wasserstein L^1 distance) by

$$\mathcal{K}(Q, Q') := \inf_{Z\sim Q, Z'\sim Q'} \tilde{E}\Delta(Z, Z'),$$

where the infimum is taken over all random variables Z and Z' with respective laws Q and Q', defined on a common probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$. Recall that the Markov kernel of the process $(Z_t)_{t\in\mathbb{Z}}$ is denoted by π_θ^Z. The following result follows immediately from Proposition 1.

Proposition 2. *Suppose that condition (A1) is fulfilled and that $\theta \in \Theta_0$. Let Q and Q' be arbitrary distributions on the state space S. Then, for $\kappa < 1$ given in Proposition 1,*

$$\mathcal{K}(Q\pi_\theta^Z, Q'\pi_\theta^Z) \leq \kappa \; \mathcal{K}(Q, Q').$$

Proposition 2 shows that the mapping π_θ^Z is contractive. Therefore, we can conclude by the Banach fixed point theorem that the Markov process $(Z_t)_{t\in\mathbb{Z}}$ has a unique stationary distribution.

Corollary 1. *Suppose that condition (A1) is fulfilled and that $\theta \in \Theta_0$. Then the Markov process $(Z_t)_{t\in\mathbb{Z}}$ with transition kernel π_θ^Z has a unique stationary distribution.*

The next lemma states that condition (A1) implies that the corresponding random variables have bounded moments of all orders. Still with a view toward our model-based bootstrap method, we have to accommodate the feature of nonstationarity and we require uniformity of the result in a neighborhood of the true parameter θ_0.

Lemma 2. *Suppose that condition (A1) is fulfilled and that $\theta \in \Theta_0$. Let $k \in \mathbb{N}$ be arbitrary.*

(i) Suppose that the process $(Z_t)_{t\in\mathbb{N}_0}$ with Markov kernel π_θ^Z is started with pre-sample values $Z_0 = (X_0, \dots, X_{1-p}, \lambda_0, \dots, \lambda_{1-q})$ such that $\max\{\mathbb{E}X_0^k, \dots, \mathbb{E}X_{1-p}^k, \mathbb{E}\lambda_0^k, \dots, \mathbb{E}\lambda_{1-q}^k\} < \infty$.

Then there exist constants $C_k, D_k < \infty$ *which only depend on* $\sum_{i=1}^{p} c_i + \sum_{j=1}^{q} d_j$ *and* $\max\{\mathbb{E}X_0^k, \ldots, \mathbb{E}X_{1-p}^k, \mathbb{E}\lambda_0^k, \ldots, \mathbb{E}\lambda_{1-q}^k\}$ *such that*

$$EX_t^k \le C_k \qquad \text{and} \qquad E\lambda_t^k \le D_k \qquad \forall t \in \mathbb{N}.$$

(ii) *If the process* $(Z_t)_{t\in\mathbb{Z}}$ *with Markov kernel* π_θ^Z *is in its stationary regime, then there exist constants* $C_k', D_k' < \infty$ *which only depend on* $\sum_{i=1}^{p} c_i + \sum_{j=1}^{q} d_j$ *such that*

$$EX_t^k \le C_k' \qquad \text{and} \qquad E\lambda_t^k \le D_k' \qquad \forall t \in \mathbb{Z}.$$

2.3 | Absolute regularity

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{A}_1$, $\mathcal{A}_2$ be two sub-σ-algebras of $\mathcal{A}$. Then the coefficient of absolute regularity is defined as

$$\beta(\mathcal{A}_1, \mathcal{A}_2) = E\left[\sup\{|P(B|\mathcal{A}_1) - P(B)| : \quad B \in \mathcal{A}_2\}\right].$$

For a process $\mathbf{Y} = (Y_t)_t$ on $(\Omega, \mathcal{A}, P)$, the coefficients of absolute regularity at the time point k are defined as

$$\beta^Y(k, n) = \beta\left(\sigma(Y_k, Y_{k-1}, \ldots), \sigma(Y_{k+n}, Y_{k+n+1}, \ldots)\right)$$

and the (global) coefficients of absolute regularity as

$$\beta^Y(n) = \sup_k \left\{\beta^Y(k, n)\right\}.$$

For the count process $(X_t)_t$ on $(\Omega, \mathcal{F}, \mathbb{P})$, we obtain the following estimate of the coefficients of absolute regularity at the point k.

$$\begin{aligned}
\beta^X(k, n) &= \beta\left(\sigma(X_k, X_{k-1}, \ldots), \sigma(X_{k+n}, X_{k+n+1}, \ldots)\right) \\
&\le \beta\left(\sigma(Z_k, Z_{k-1}, \ldots), \sigma(X_{k+n}, X_{k+n+1}, \ldots)\right) \\
&= \beta\left(\sigma(Z_k), \sigma(X_{k+n}, X_{k+n+1}, \ldots)\right) \\
&= \mathbb{E}\left[\sup_{C\in\sigma(\mathcal{Z})}\{|\mathbb{P}_\theta\left((X_{k+n}, X_{k+n+1}, \ldots) \in C|Z_k\right) - \mathbb{P}_\theta\left((X_{k+n}, X_{k+n+1}, \ldots) \in C\right)|\}\right],
\end{aligned} \tag{4}$$

where $\mathcal{Z} = \{A \times \mathbb{Z} \times \mathbb{Z} \times \ldots \mid A \subseteq \mathbb{N}_0^m, m \in \mathbb{N}\}$ is the system of cylinder sets. At this point we employ a coupling argument. Let $((\tilde{Z}_t, \tilde{Z}_t'))_{t\in\mathbb{N}_0}$ be a Markov chain on a probability space $(\tilde{\Omega}, \widetilde{\mathcal{F}}, \tilde{P})$ with transition kernel $\tilde{\pi}_\theta^Z$ and independent variables $\tilde{Z}_k, \tilde{Z}_k' \sim \mathbb{P}_\theta^{Z_k}$. Then

$$\begin{aligned}
&\mathbb{E}\left[\sup_{C\in\sigma(\mathcal{Z})}\{|\mathbb{P}_\theta\left((X_{k+n}, X_{k+n+1}, \ldots) \in C|Z_k\right) - \mathbb{P}_\theta\left((X_{k+n}, X_{k+n+1}, \ldots) \in C\right)|\}\right] \\
&\le \tilde{E}\left[\sup_{C\in\sigma(\mathcal{Z})}\left\{\left|\tilde{P}\left((\tilde{X}_{k+n}, \tilde{X}_{k+n+1}, \ldots) \in C|\tilde{Z}_k\right) - \tilde{P}\left((\tilde{X}_{k+n}', \tilde{X}_{k+n+1}', \ldots) \in C|\tilde{Z}_k'\right)\right|\right\}\right] \\
&\le \tilde{P}\left(\tilde{X}_{n+k+l} \neq \tilde{X}_{n+k+l}' \quad \text{for some } \ l \ge 0\right) \\
&\le \sum_{l=0}^{\infty} \tilde{P}\left(\tilde{X}_{n+k+l} \neq \tilde{X}_{n+k+l}'\right).
\end{aligned} \tag{5}$$

Since $\tilde{X}_{n+k+l}$ and $\tilde{X}'_{n+k+l}$ are integer-valued we obtain that

$$\tilde{P}\left(\tilde{X}_{n+k+l} \neq \tilde{X}'_{n+k+l}\right) \leq \frac{1}{\gamma_1} \tilde{E}\Delta(\tilde{Z}_{n+k+l}, \tilde{Z}'_{n+k+l}). \tag{6}$$

Furthermore, by Proposition 1

$$\tilde{E}\Delta(\tilde{Z}_{n+k+l}, \tilde{Z}'_{n+k+l}) \leq \kappa^{n+l} \tilde{E}\Delta(\tilde{Z}_k, \tilde{Z}'_k), \tag{7}$$

where $\sup_k \left\{\tilde{E}\Delta(\tilde{Z}_k, \tilde{Z}'_k)\right\} < \infty$. From (4) to (7) we obtain absolute regularity of the count process $(X_t)_{t\in\mathbb{Z}}$ with exponentially decaying coefficients.

Theorem 1. *Suppose that condition (A1) is fulfilled and that $\theta \in \Theta_0$. Suppose either that the process $(X_t)_{t\in\mathbb{Z}}$ is in its stationary regime or that the process $(X_t)_{t\in\mathbb{N}_0}$ is started with presample values $Z_0 = (X_0, \dots, X_{1-p}, \lambda_0, \dots, \lambda_{1-q})$ such that* $\max\{\mathbb{E}X_0^k, \dots, \mathbb{E}X_{1-p}^k, \mathbb{E}\lambda_0^k, \dots, \mathbb{E}\lambda_{1-q}^k\} < \infty$. *Then there exists some $\rho < 1$ such that*

$$\beta^X(n) = O\left(\rho^n\right).$$

3 | BOOTSTRAP

We assume that $X_1, \dots, X_n$ are observed, where $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary and satisfies (A1). Let $\hat{\theta}_n$ be any consistent estimator of the true parameter θ_0. A typical example is the conditional maximum likelihood estimator investigated, for example, in Fokianos and Tjostheim (2011); Fokianos and Tjostheim (2012) or, in the special case of an integer-valued ARCH model, a least squares estimator. The bootstrap process is generated as follows. To initiate the process, we choose presample values $X_0^*, \dots, X_{1-p}^*, \lambda_0^*, \dots, \lambda_{1-q}^*$. The next value of the intensity process has to obey the model equation (1b), that is,

$$\lambda_1^* = f_{\hat{\theta}_n}(X_0^*, \dots, X_{1-p}^*, \lambda_0^*, \dots, \lambda_{1-q}^*).$$

Then, conditioned on the past values, X_1^* has to follow a Poisson(λ_1^*) distribution. This process can be repeated arbitrarily often. If $X_{t-1}^*, \dots, X_{1-p}^*$ and $\lambda_{t-1}^*, \dots, \lambda_{1-q}^*$ are generated, we choose the next values such that

$$\lambda_t^* = f_{\hat{\theta}_n}(X_{t-1}^*, \dots, X_{t-p}^*, \lambda_{t-1}^*, \dots, \lambda_{t-q}^*),$$

and, again conditioned on the past,

$$X_t^* \sim \text{Pois}(\lambda_t^*).$$

3.1 | Coupling

Bootstrap methods are typically used for the construction of confidence regions for parameters or for determining critical values of tests. To ensure a versatile applicability it is necessary that the bootstrap process, conditioned on the original sample, mimics the stochastic behavior of the original process as good as possible. This similarity can be shown in a most transparent way by a coupling of the original sample $(X_1, \lambda_1), \dots, (X_n, \lambda_n)$ and its bootstrap counterpart $(X_1^*, \lambda_1^*), \dots, (X_n^*, \lambda_n^*)$. Somewhat surprisingly, so far this natural approach was extremely rarely used in statistics. Using Mallows metric to measure the distance between variables from the original and the bootstrap process, it was implicitly employed in the context of independent random variables by Bickel and Freedman (1981) and Freedman (1981). A more explicit use of coupling was made, in the context of U- and V-statistics, but again in the independent case, by Dehling and Mikosch (1994) and Leucht and Neumann (2009). For dependent data, this approach was adopted by Leucht and Neumann (2013) and Leucht, Neumann, and Kreiss (2015).

To display the required similarity of the original process $((X_t, \lambda_t))_{t=1,\dots,n}$ and its bootstrap counterpart $((X_t^*, \lambda_t^*))_{t=1,\dots,n}$ we employ a coupling of the corresponding Markov processes $(Z_t)_{t\in\mathbb{N}_0}$ and $(Z_t^*)_{t\in\mathbb{N}_0}$. That is, we construct on a suitable probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ copies $(\tilde{Z}_t)_{t\in\mathbb{N}_0}$ and $(\tilde{Z}_t^*)_{t\in\mathbb{N}_0}$ such that

$$\tilde{P}^{\tilde{Z}_t|\tilde{Z}_{t-1}=z} = \mathbb{P}_{\theta_0}^{Z_t|Z_{t-1}=z} \quad \text{and} \quad \tilde{P}^{\tilde{Z}_t^*|\tilde{Z}_{t-1}^*=z} = \mathbb{P}_{\hat{\theta}_n}^{Z_t^*|Z_{t-1}^*=z}.$$

With this coupling, we have to take care that $\tilde{Z}_t$ and $\tilde{Z}_t^*$ are close to each other. We denote the corresponding Markov kernel which transfers $(\tilde{Z}_{t-1}, \tilde{Z}_{t-1}^*)$ into $(\tilde{Z}_t, \tilde{Z}_t^*)$ by π^{Z,Z^*}. Furthermore, in order to derive mixing properties of the coupled process $((\tilde{Z}_t, \tilde{Z}_t^*))_{t\in\mathbb{N}_0}$, we show that the Markov kernel is contractive. As in the previous Section 2, this is achieved by a further coupling of two versions, $((\tilde{Z}_t, \tilde{Z}_t^*))_{t\in\mathbb{N}_0}$ and $((\tilde{Z}_t', \tilde{Z}_t^{*\prime}))_{t\in\mathbb{N}_0}$ of the above process. This Markov kernel will be denoted $\tilde{\pi}^{Z,Z^*}$. Suppose that $\tilde{Z}_{t-1} = z := (x_1, \dots, x_p, \lambda_1, \dots, \lambda_q)$, $\tilde{Z}_{t-1}' = z' := (x_1', \dots, x_p', \lambda_1', \dots, \lambda_q')$, $\tilde{Z}_{t-1}^* = z^* := (x_1^*, \dots, x_p^*, \lambda_1^*, \dots, \lambda_q^*)$, and $\tilde{Z}_{t-1}^{*\prime} = z^{*\prime} := (x_1^{*\prime}, \dots, x_p^{*\prime}, \lambda_1^{*\prime}, \dots, \lambda_q^{*\prime})$ are given. Let $(N_t(u))_{u\geq 0}$ be a Poisson process with unit intensity which is independent of these random variables. Then we set $\tilde{Z}_t = (\tilde{X}, x_1, \dots, x_{p-1}, \tilde{\lambda}, \lambda_1, \dots, \lambda_{q-1})$, $\tilde{Z}_t' = (\tilde{X}', x_1', \dots, x_{p-1}', \tilde{\lambda}', \lambda_1', \dots, \lambda_{q-1}')$, $\tilde{Z}_t^* = (\tilde{X}^*, x_1^*, \dots, x_{p-1}^*, \tilde{\lambda}^*, \lambda_1^*, \dots, \lambda_{q-1}^*)$, and $\tilde{Z}_t^{*\prime} = (\tilde{X}^{*\prime}, x_1^{*\prime}, \dots, x_{p-1}^{*\prime}, \tilde{\lambda}^{*\prime}, \lambda_1^{*\prime}, \dots, \lambda_{q-1}^{*\prime})$. Here, according to the model equation, $\tilde{\lambda} = f_{\theta_0}(z)$, $\tilde{\lambda}' = f_{\theta_0}(z')$, $\tilde{\lambda}^* = f_{\hat{\theta}_n}(z^*)$, and $\tilde{\lambda}^{*\prime} = f_{\hat{\theta}_n}(z^{*\prime})$. We obtain the desired proximity by generating $\tilde{X}, \tilde{X}', \tilde{X}^*$, and $\tilde{X}^{*\prime}$ from the same standard Poisson process $(N_t(u))_{u\geq 0}$, that is, $\tilde{X} = N_t(\tilde{\lambda})$, $\tilde{X}' = N_t(\tilde{\lambda}')$, $\tilde{X}^* = N_t(\tilde{\lambda}^*)$, and $\tilde{X}^{*\prime} = N_t(\tilde{\lambda}^{*\prime})$.

To prove consistency of the bootstrap approximation, we impose the following conditions.

(A2) $\hat{\theta}_n \xrightarrow{P} \theta_0$ and there exists some $\delta > 0$ such that $\{\theta \in \Theta : \|\theta - \theta_0\| \leq \delta\} \subseteq \Theta_0$.

(A3) There exists some $M > 0$ such that

$$\left|f_\theta(x_1, \dots, x_p, \lambda_1, \dots, \lambda_q) - f_{\theta_0}(x_1, \dots, x_p, \lambda_1, \dots, \lambda_q)\right| \leq M \|\theta - \theta_0\| \left(\sum_{i=1}^p x_i + \sum_{j=1}^q \lambda_j \right),$$

holds for all $(x_1, \dots, x_p, \lambda_1, \dots, \lambda_q) \in S$ and all $\theta \in \Theta_0$.

Remark 1. $\hat{\theta}_n \xrightarrow{P} \theta_0$ implies that there exist null sequences $(\delta_n)_{n\in\mathbb{N}}$ and $(\rho_n)_{n\in\mathbb{N}}$ such that

$$\mathbb{P}_{\theta_0}\left(\|\hat{\theta}_n - \theta_0\| > \delta_n \right) \leq \rho_n. \tag{8}$$

Hence, (A2) implies that

$$\mathbb{P}_{\theta_0}\left(\hat{\theta}_n \in \Theta_0 \right) \underset{n\to\infty}{\longrightarrow} 1.$$

3.2 | Stationarity and absolute regularity of the coupled process

Note that the uniform contraction condition (A1) is fulfilled for $f_{\hat{\theta}_n}$ if $\hat{\theta}_n \in \Theta_0$. We remind the reader that we denote the Markov kernel of the process $((\tilde{Z}_t, \tilde{Z}_t^*))_{t\in\mathbb{N}_0}$ by π^{Z,Z^*} and that of $((\tilde{Z}_t, \tilde{Z}_t^*), (\tilde{Z}_t', \tilde{Z}_t^{*\prime}))_{t\in\mathbb{N}_0}$ by $\tilde{\pi}^{Z,Z^*}$. In case of $\hat{\theta}_n \in \Theta_0$, we obtain by Proposition 1(ii) that

$$\tilde{E}\left[\Delta(\tilde{Z}_t, \tilde{Z}_t') + \Delta(\tilde{Z}_t^*, \tilde{Z}_t^{*\prime})\right] \leq \kappa\, \tilde{E}\left[\Delta(\tilde{Z}_{t-1}, \tilde{Z}_{t-1}') + \Delta(\tilde{Z}_{t-1}^*, \tilde{Z}_{t-1}^{*\prime})\right]. \tag{9}$$

This means that the Markov kernel $\tilde{\pi}^{Z,Z^*}$ is contractive w.r.t. the Kantorovich distance $\mathcal{K}^{(2)}$, where

$$\mathcal{K}^{(2)}(Q, Q') = \inf_{(Z_1,Z_2)\sim Q,(Z_1',Z_2')\sim Q'} E\left[\Delta(Z_1, Z_1') + \Delta(Z_2, Z_2')\right].$$

We obtain, analogously to Proposition 2, that

$$\mathcal{K}^{(2)}(Q\pi^{Z,Z^*}, Q'\pi^{Z,Z^*}) \leq \kappa\, \mathcal{K}^{(2)}(Q, Q'), \tag{10}$$

holds for arbitrary distributions Q and Q' on the state space $S \times S$. This leads immediately to the following theorem.

Theorem 2. *Suppose that (A1), (A2), and (A3) are fulfilled. Then the following statements hold true with a probability tending to one, as $n \to \infty$.*

(i) The process $((\tilde{Z}_t, \tilde{Z}_t^*))_t$ has a unique stationary distribution.

(ii) Suppose either that the process $((\tilde{Z}_t, \tilde{Z}_t^*))_{t\in\mathbb{Z}}$ is in its stationary regime or that the process $((\tilde{Z}_t, \tilde{Z}_t^*))_{t\in\mathbb{N}_0}$ is started with pre-sample values $\tilde{Z}_0 = (\tilde{X}_0, \dots, \tilde{X}_{1-p}, \tilde{\lambda}_0, \dots, \tilde{\lambda}_{1-q})$ and $\tilde{Z}_0^* = (\tilde{X}_0^*, \dots, \tilde{X}_{1-p}^*, \tilde{\lambda}_0^*, \dots, \tilde{\lambda}_{1-q}^*)$ such that $\max\{\tilde{E}\tilde{X}_0^k, \dots, \tilde{E}\tilde{X}_{1-p}^k, \tilde{E}\tilde{\lambda}_0^k, \dots, \tilde{E}\tilde{\lambda}_{1-q}^k\} \leq C$ and $\max\{\tilde{E}(\tilde{X}_0^*)^k, \dots, (\tilde{E}\tilde{X}_{1-p}^*)^k, \tilde{E}(\tilde{\lambda}_0^*)^k, \dots, \tilde{E}(\tilde{\lambda}_{1-q}^*)^k\} \leq C$, for some $C < \infty$. Then there exists some $\rho < 1$ such that

$$\beta^{(\tilde{X},\tilde{X}^*)}(n) = O(\rho^n).$$

At this point we recall once more that any consistent estimator $\hat{\theta}_n$ of θ_0 falls into Θ_0 with a probability tending to one. This means that the above regularity properties of the coupled process are fulfilled asymptotically.

3.3 | Proximity of the original and the bootstrap process

Suppose now that (A1), (A2), and (A3) are fulfilled and that the process $(Z_t)_{t\in\mathbb{Z}}$ is in its stationary regime. We derive a recursion for the process $((\tilde{Z}_t, \tilde{Z}_t^*))_{t\in\mathbb{N}_0}$. We have that

$$\begin{aligned}
\tilde{E}\left|\tilde{X}_t^* - \tilde{X}_t\right| &= \tilde{E}\left[\tilde{E}\left(|\tilde{X}_t^* - \tilde{X}_t| \,\Big|\, \tilde{\lambda}_t^*, \tilde{\lambda}_t\right)\right] = \tilde{E}\left|\tilde{\lambda}_t^* - \tilde{\lambda}_t\right| \\
&\leq \tilde{E}\left|f_{\hat{\theta}_n}(\tilde{Z}_{t-1}^*) - f_{\hat{\theta}_n}(\tilde{Z}_{t-1})\right| + \tilde{E}\left|f_{\hat{\theta}_n}(\tilde{Z}_{t-1}) - f_{\theta_0}(\tilde{Z}_{t-1})\right| \\
&\leq \sum_{i=1}^{p} c_i\, \tilde{E}\left|\tilde{X}_{t-i}^* - \tilde{X}_{t-i}\right| + \sum_{j=1}^{q} d_j\, \tilde{E}\left|\tilde{\lambda}_{t-j}^* - \tilde{\lambda}_{t-j}\right| + M\, \|\hat{\theta}_n - \theta_0\|\, \tilde{E}\|\tilde{Z}_{t-1}\|_1.
\end{aligned}$$

It follows from Lemma 1 and the previous display that

$$
\begin{aligned}
&\tilde{E}\Delta(\tilde{Z}_t^*, \tilde{Z}_t) \\
&= \gamma_1\, \tilde{E}\left|\tilde{X}_t^* - \tilde{X}_t\right| + \sum_{i=2}^{p} \gamma_i\, \tilde{E}\left|\tilde{X}_{t-i+1}^* - \tilde{X}_{t-i+1}\right| + \delta_1\, \tilde{E}\left|\tilde{\lambda}_t^* - \tilde{\lambda}_t\right| + \sum_{j=2}^{q} \gamma_j\, \tilde{E}\left|\tilde{\lambda}_{t-j+1}^* - \tilde{\lambda}_{t-j+1}\right| \\
&\leq \kappa\, \tilde{E}\Delta\left(\tilde{Z}_{t-1}^*, \tilde{Z}_{t-1}\right) + (\gamma_1 + \delta_1) M\, \|\hat{\theta}_n - \theta_0\|\, \tilde{E}\|\tilde{Z}_{t-1}\|_1 \\
&\leq \kappa^2\, \tilde{E}\Delta\left(\tilde{Z}_{t-2}^*, \tilde{Z}_{t-2}\right) + \kappa(\gamma_1 + \delta_1) M\, \|\hat{\theta}_n - \theta_0\|\, \tilde{E}\|\tilde{Z}_{t-2}\|_1 + (\gamma_1 + \delta_1) M\, \|\hat{\theta}_n - \theta_0\|\, \tilde{E}\|\tilde{Z}_{t-1}\|_1 \\
&\leq \ldots \leq \left(\sum_{s=0}^{t-1} \kappa^s\right) M\, \mathbb{E}\|Z_0\|_1 (\gamma_1 + \delta_1) \|\hat{\theta}_n - \theta_0\| + \kappa^t\, \tilde{E}\Delta\left(\tilde{Z}_0^*, \tilde{Z}_0\right).
\end{aligned}
$$

This leads to the following theorem.

Theorem 3. *Suppose that (A1), (A2), and (A3) are fulfilled and that $(Z_t)_{t\in\mathbb{Z}}$ is in its stationary regime. Then*

(i) If $(Z_t^*)_{t\in\mathbb{N}_0}$ is started with a pre-sample value Z_0^*, then

$$
\tilde{E}\Delta\left(\tilde{Z}_t^*, \tilde{Z}_t\right) \;\leq\; \frac{M\, \mathbb{E}\|Z_0\|_1 (\gamma_1 + \delta_1)}{1 - \kappa}\, \|\hat{\theta}_n - \theta_0\| + \kappa^t\, \tilde{E}\Delta\left(\tilde{Z}_0^*, \tilde{Z}_0\right).
$$

(ii) If $(Z_t^*)_{t\in\mathbb{N}_0}$ is in its stationary regime, then

$$
\tilde{E}\Delta\left(\tilde{Z}_t^*, \tilde{Z}_t\right) \;\leq\; \frac{M\, \mathbb{E}\|Z_0\|_1 (\gamma_1 + \delta_1)}{1 - \kappa}\, \|\hat{\theta}_n - \theta_0\|.
$$

4 | BOOTSTRAP CONSISTENCY FOR STATISTICS OF DIFFERENT TYPE

In this section we exhibit some examples of important statistics for which consistency of their bootstrap approximation can be easily shown on the basis of the coupling results given in Theorems 2 and 3. Let $S_n = h_n(X_1, \ldots, X_n; \theta_0)$ be a statistic of interest, for example, a test statistic or a statistic used as a starting point for the construction of a confidence interval. Typically, we can employ standard tools to show that $(S_n)_{n\in\mathbb{N}}$ converges in distribution to some random variable, say Y. The distribution of the latter depends often on one or more unknown parameters which suggests the application of a bootstrap approximation. Let $S_n^* = h_n(X_1^*, \ldots, X_n^*; \hat{\theta}_n)$ be the bootstrap counterpart of S_n. We denote by $\tilde{S}_n = h_n(\tilde{X}_1, \ldots, \tilde{X}_n; \theta_0)$ and $\tilde{S}_n^* = h_n(\tilde{X}_1^*, \ldots, \tilde{X}_n^*; \hat{\theta}_n)$ the coupled versions of S_n and S_n^*, respectively, where the coupling of the underlying random variables is described in the previous section. Below we show for some examples that the approximation result for $\tilde{X}_t$ and $\tilde{X}_t^*$ given in Theorem 3 immediately implies, in conjunction with the mixing result in Theorem 2, that

$$
\left|\tilde{S}_n - \tilde{S}_n^*\right| \xrightarrow{\tilde{P}} 0. \tag{11}
$$

For the purpose of determining critical values of tests or for establishing confidence intervals we need, however, convergence of the respective cumulative distribution functions. Let F_{S_n} be the cdf of the statistic S_n and $F_{S_n^*}$ be that of the conditional distribution of S_n^* given $X_1, \dots, X_n$. If in addition Y has a continuous distribution, then (11) implies by lemma 2.11 of van der Vaart (1998) that

$$\sup_x \left| F_{S_n}(x) - F_{S_n^*}(x) \right| \stackrel{P}{\to} 0. \tag{12}$$

This ensures that a test based on S_n with a critical value chosen by the corresponding quantile of $F_{S_n^*}$ reaches asymptotically the prescribed size.

In what follows we assume that (A1), (A2), and (A3) are fulfilled which means that we have the results of Theorems 2 and 3 at our disposal. Recall from Remark 1 that it follows from (A2) that there exist null sequences $(\delta_n)_{n\in\mathbb{N}}$ and $(\rho_n)_{n\in\mathbb{N}}$ such that

$$\mathbb{P}_{\theta_0}\left(\|\hat{\theta}_n - \theta_0\| > \delta_n \right) \ \le \ \rho_n,$$

which implies in particular that

$$\mathbb{P}_{\theta_0}\left(\hat{\theta}_n \in \Theta_0 \right) \underset{n\to\infty}{\to} 1.$$

We consider some statistics and argue that our approximation results imply without much additional effort consistency of the corresponding bootstrap approximations.

Example 1. Sample mean

Suppose we observe $X_1, \dots, X_n$, where the underlying process $(Z_t)_{t\in\mathbb{Z}}$ is strictly stationary. We focus on the mean $\mu = E_{\theta_0}X_1$, which is estimated by its sample version, $\overline{X}_n = n^{-1}\sum_{t=1}^n X_t$. It follows from a central limit theorem for absolutely regular random variables that

$$S_n \ := \sqrt{n}\left(\overline{X}_n - \mu \right) \stackrel{d}{\to} Y \ \sim \ N(0, \sigma_\infty^2), \tag{13}$$

where $\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} cov(X_k, X_0)$. In order to construct a confidence interval for μ with an asymptotic coverage probability of $\gamma \in (0,1)$, we can either approximate σ_∞^2 by a consistent estimator and rely on the normal approximation or use a bootstrap approximation to the distribution of S_n. Such an approximation is given by

$$S_n^* = \sqrt{n}\left(\overline{X}_n^* - E^*\overline{X}_n^* \right),$$

where $\overline{X}_n^* = n^{-1}\sum_{t=1}^n X_t^*$ and $E^*\overline{X}_n^* = E(\overline{X}_n^* | X_1, \dots, X_n)$ is the expectation in the "bootstrap world." In order to show that this is asymptotically correct, we could employ at this point a suitable central limit theorem for triangular arrays of dependent random variables and prove that

$$S_n^* \stackrel{d}{\to} Y \qquad \text{in probability.}$$

On the other hand, as explained in what follows, we could simply use the results of Theorems 2 and 3 and obtain the desired consistency almost effortlessly.

We assume that (A1), (A2), and (A3) are fulfilled. We suppose that the bootstrap process is started with pre-sample values $Z_0^* = (X_0^*, \dots, X_{1-p}^*, \lambda_0^*, \dots, , \lambda_{1-q}^*)$ such that $E_{\theta_0}E^*\left[\sum_{i=1}^p X_{1-i}^* + \sum_{j=1}^q \lambda_{1-j}^*\right] < \infty$. We assume from here on that

$$\|\hat{\theta}_n - \theta_0\| \;\leq\; \delta_n \qquad \text{and} \qquad \hat{\theta}_n \in \Theta_0,$$

which happens with a probability tending to 1 and which implies that the bootstrap process behaves as desired. Theorem 3 yields, for the coupled versions of the original and bootstrap random variables,

$$\tilde{E}\left|\tilde{X}_t - \tilde{X}_t^*\right| = O\left(\delta_n \vee \kappa^t\right). \tag{14}$$

Since moments of arbitrary order of these random variables are bounded we obtain, for all $k \in \mathbb{N}$, arbitrary $\alpha \in (0, 1)$ and $\beta = 1/(1 - \alpha)$, that

$$\begin{aligned}
\tilde{E}\left[\left|\tilde{X}_t - \tilde{X}_t^*\right|^k\right] &= \tilde{E}\left[\left|\tilde{X}_t - \tilde{X}_t^*\right|^\alpha \left|\tilde{X}_t - \tilde{X}_t^*\right|^{k-\alpha}\right] \\
&\leq \left(\tilde{E}\left|\tilde{X}_t - \tilde{X}_t^*\right|\right)^\alpha \left(\tilde{E}\left[\left|\tilde{X}_t - \tilde{X}_t^*\right|^{(k-\alpha)\beta}\right]\right)^{1/\beta} \\
&= O\left(\delta_n^\alpha \vee \kappa^{\alpha t}\right).
\end{aligned} \tag{15}$$

We obtain in particular

$$\tilde{E}\left[\left|(\tilde{X}_t - \mu) - (\tilde{X}_t^* - E^*X_t^*)\right|^2\right] \;\leq\; \tilde{E}\left[\left|\tilde{X}_t - \tilde{X}_t^*\right|^2\right] = O\left(\delta_n^\alpha \vee \kappa^{\alpha t}\right).$$

Since absolute regularity implies strong mixing we can use a covariance inequality for strongly mixing processes and obtain that

$$\begin{aligned}
& cov\left(\left((\tilde{X}_s - \mu) - (\tilde{X}_s^* - E^*X_s^*)\right), \left((\tilde{X}_t - \mu) - (\tilde{X}_t^* - E^*X_n^*)\right)\right) \\
&= O\left(\rho^{|s-t|/3} \left(\tilde{E}\left|(\tilde{X}_s - \mu) - (\tilde{X}_s^* - E^*X_s^*)\right|^3\right)^{1/3} \left(\tilde{E}\left|(\tilde{X}_t - \mu) - (\tilde{X}_t^* - E^*X_t^*)\right|^3\right)^{1/3}\right) \\
&= O\left(\rho^{|s-t|/3} \left(\delta_n^{\alpha/3} \vee \kappa^{\alpha s/3}\right) \left(\delta_n^{\alpha/3} \vee \kappa^{\alpha t/3}\right)\right).
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
\tilde{E}\left[\left|\tilde{S}_n - \tilde{S}_n^*\right|^2\right] &= \frac{1}{n} \sum_{s,t=1}^{n} cov\left(\left((\tilde{X}_s - \mu) - (\tilde{X}_s^* - E^*X_s^*)\right), \left((\tilde{X}_t - \mu) - (\tilde{X}_t^* - E^*X_n^*)\right)\right) \\
&= \frac{1}{n} \sum_{s,t=1}^{[\log(\delta_n)/\log(\kappa)]} O\left(\rho^{|s-t|/3}\, \kappa^{\alpha s/3}\, \kappa^{\alpha t/3}\right) \\
&+ \frac{2}{n} \sum_{s=1}^{[\log(\delta_n)/\log(\kappa)]} \sum_{t=[\log(\delta_n)/\log(\kappa)]+1}^{n} O\left(\rho^{|s-t|/3}\, \kappa^{\alpha s/3}\, \delta_n^{\alpha/3}\right) \\
&+ \frac{1}{n} \sum_{s,t=[\log(\delta_n)/\log(\kappa)]+1}^{n} O\left(\rho^{|s-t|/3}\, \delta_n^{2\alpha/3}\right) \\
&= o(1).
\end{aligned} \tag{16}$$

that is, (11) is fulfilled.

Example 2. The above results can be extended to so-called generalized means as considered in Jentsch and Weiss (2019), in the context of integer-valued autoregressive (INAR) processes. These authors considered statistics T_n of the form

$$T_n = \sqrt{n}\left[f\left(\frac{1}{n-m+1}\sum_{t=1}^{n-m+1} g\left(X_t, \ldots, X_{t+m-1}\right)\right) - f\left(E_{\theta_0} g\left(X_1, \ldots, X_m\right)\right)\right],$$

for sufficiently regular functions $g : \mathbb{R}^m \to \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$. Simple examples of such functions are given by $g_1(x) = \mathbb{1}(x = k)$, $g_2(x) = \mathbb{1}(x \le k)$, $k \in \mathbb{N}_0$, and $f(y) = y$. This leads to the statistics

$$T_n^{(1)} = \sqrt{n}\left[n^{-1}\sum_{t=1}^{n} \mathbb{1}\left(X_t = k\right) - P_{\theta_0}\left(X_1 = k\right)\right]$$

and

$$T_n^{(2)} = \sqrt{n}\left[n^{-1}\sum_{t=1}^{n} \mathbb{1}\left(X_t \le k\right) - P_{\theta_0}\left(X_1 \le k\right)\right],$$

respectively. Since $\tilde{X}_t$ and $\tilde{X}_t^*$ are integer-valued we have $\left|g_i(\tilde{X}_t) - g_i(\tilde{X}_t^*)\right| \le |\tilde{X}_t - \tilde{X}_t^*|$, for $i = 1, 2$. Hence, approximation (14) immediately yields that

$$\tilde{E}\left|g_i\left(\tilde{X}_t\right) - g_i\left(\tilde{X}_t^*\right)\right| = O\left(\delta_n \vee \kappa^t\right),$$

and we can use the same arguments as above to show consistency of the corresponding bootstrap approximations. Some statistics such as, for example,

$$T_n^{(3)} = \sqrt{n}\left[\frac{\sum_{t=1}^{n-1} \mathbb{1}((X_{t+1}, X_t) = (i,j))}{\sum_{t=1}^{n-1} \mathbb{1}(X_t = j)} - P_{\theta_0}\left(X_2 = j | X_1 = i\right)\right]$$

can probably treated similarly, however, this requires a few additional arguments. In the related context of INAR processes, Jentsch and Weiß (2019) provided a more general result on bootstrap consistency, under some sort of high-level assumptions.

Example 3. Autocovariances

Suppose again that $(Z_t)_{t\in\mathbb{Z}}$ is strictly stationary, $\mu = E_{\theta_0} X_1$. Let $\gamma : \mathbb{Z} \to \mathbb{R}$ with $\gamma(k) = cov(X_k, X_0)$ be the autocovariance function of the count process. Based on observations $X_1, \ldots, X_n$, a natural estimator of $\gamma(k)$ is given by

$$\hat{\gamma}_n(k) = \frac{1}{n}\sum_{t=1}^{n-|k|} (X_{t+|k|} - \overline{X}_n)(X_t - \overline{X}_n).$$

Since $E_{\theta_0}[(\overline{X}_n - \mu)^2] = O(n^{-1})$ we obtain that

$$T_n := \sqrt{n}\left(\hat{\gamma}_n(k) - \gamma(k)\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-|k|}\left[(X_{t+|k|} - \mu)(X_t - \mu) - \gamma(k)\right] + O_P\left(n^{-1/2}\right).$$

Using this we obtain by a central limit theorem for dependent random variables that

$$T_n \xrightarrow{d} Y_2 \sim N(0, \tau_\infty^2), \tag{17}$$

where $\tau_\infty^2 = \sum_{l=-\infty}^{\infty} cov\left((X_{|k|+l} - \mu)(X_l - \mu), (X_{|k|} - \mu)(X_0 - \mu)\right)$. A bootstrap version of T_n is given by

$$T_n^* = \sqrt{n}\left(\hat{\gamma}_n^*(k) - E^*\hat{\gamma}_n^*(k)\right),$$

where

$$\hat{\gamma}_n^*(k) = \frac{1}{n}\sum_{t=1}^{n-|k|}(X_{t+|k|}^* - \overline{X}_n^*)(X_t^* - \overline{X}_n^*).$$

As above, we denote by $\tilde{X}_t$ and $\tilde{X}_t^*$ the coupled versions of X_t and X_t^*, respectively. We can show analogously to (15) that, for $\alpha \in (0,1)$,

$$\tilde{E}\left[\left((\tilde{X}_{t+|k|} - \bar{\tilde{X}}_n)(\tilde{X}_t - \bar{\tilde{X}}_n) - (\tilde{X}_{t+|k|}^* - \bar{\tilde{X}}_n^*)(\tilde{X}_t^* - \bar{\tilde{X}}_n^*)\right)^k\right] = O\left(\delta_n^\alpha \vee \kappa^{\alpha t}\right). \tag{18}$$

Indeed, splitting up

$$\begin{aligned}(\tilde{X}_{t+|k|} - \bar{\tilde{X}}_n)(\tilde{X}_t - \bar{\tilde{X}}_n) - (\tilde{X}_{t+|k|}^* - \bar{\tilde{X}}_n^*)(\tilde{X}_t^* - \bar{\tilde{X}}_n^*) &= \left(\tilde{X}_{t+|k|}\tilde{X}_t - \tilde{X}_{t+|k|}^*\tilde{X}_t^*\right) - \left(\tilde{X}_t\bar{\tilde{X}}_n - \tilde{X}_t^*\bar{\tilde{X}}_n^*\right)\\ &\quad - \left(\tilde{X}_{t+|k|}\bar{\tilde{X}}_n - \tilde{X}_{t+|k|}^*\bar{\tilde{X}}_n^*\right) + \left(\bar{\tilde{X}}_n^{\,2} - \bar{\tilde{X}}_n^{*2}\right),\end{aligned}$$

and using, for example, for the first term on the right-hand side the upper estimate

$$\left|\tilde{X}_{t+|k|}\tilde{X}_t - \tilde{X}_{t+|k|}^*\tilde{X}_t^*\right| \leq \left|\tilde{X}_t - \tilde{X}_t^*\right| \tilde{X}_{t+|k|} + \left|\tilde{X}_{t+|k|} - \tilde{X}_{t+|k|}^*\right| \tilde{X}_t^*$$

we obtain from (15) and boundedness of all moments of the involved random variables that (18) holds true. Now we obtain in analogy to (16) that

$$\tilde{E}\left|\tilde{T}_n - \tilde{T}_n^*\right| = o(1), \tag{19}$$

and therefore $\left|\tilde{T}_n - \tilde{T}_n^*\right| \xrightarrow{\tilde{P}} 0$.

Example 4. Degenerate von Mises and U-statistics
As another test bed for our approximation results, we consider a bootstrap approximation for degenerate von Mises (V-)statistics. For random variables $X_1, \ldots, X_n$, a V-statistic has the form

$$V_n = \sum_{s,t=1}^{n} h(X_s, X_t),$$

where h is the so-called kernel of this statistic. A U-statistic is obtained by dropping the diagonal terms, that is,

$$U_n = \sum_{1 \le s,t \le n,\ s \ne t} h(X_s, X_t).$$

Test statistics of Cramér–von Mises type can often be approximated by V-statistics, in some special cases they have even exactly such a structure. Under the null hypothesis, they show typically a degenerate behavior, that is,

$$Eh(x, X_t) = 0 \qquad \forall x.$$

Leucht and Neumann (2013) provided a general result on bootstrap consistency for U- and V-statistics. The asymptotic behavior of these statistics was derived under the following condition.

(B1) (i) $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process with values in $\mathbb{R}^d$.
(ii) $h\,:\ \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a symmetric, continuous and non-negative definite function, that is, $\forall\ c_1, \ldots, c_m \in \mathbb{R}, x_1, \ldots, x_m \in \mathbb{R}^d$ and $m \in \mathbb{N}$, $\sum_{i,j=1}^m c_i\ c_j\ h(x_i, x_j) \ge 0$.
(iii) $Eh(X_0, X_0) < \infty$.
(iv) $E(h(x, X_t) \mid X_1, \ldots, X_{t-1}) = 0\ a.s.$ for all $x \in supp(P^{X_0})$.

Leucht and Neumann (2013, Theorem 2.1) show, under the assumption (B1), that

$$V_n \xrightarrow{d} Y := \sum_k \lambda_k\, Y_k^2 \qquad \text{and} \qquad U_n \xrightarrow{d} Y - Eh(X_0, X_0),$$

as n tends to infinity. Here, $(Y_k)_k$ is a sequence of independent standard normal random variables and $(\lambda_k)_k$ denotes the sequence of nonzero eigenvalues of the equation

$$E[h(x, X_0)\Phi(X_0)] = \lambda\ \Phi(x),$$

enumerated according to their multiplicity.

Bootstrap counterparts of these statistics are most naturally given by

$$U_n^* = \frac{1}{n} \sum_{1 \le s,t \le n,\ s \ne t} h^*(X_s^*, X_t^*) \qquad \text{and} \qquad V_n^* = \frac{1}{n} \sum_{s,t=1}^n h^*(X_s^*, X_t^*).$$

For the proof of bootstrap consistency, Leucht and Neumann (2013) imposed the following condition.

(B2) (i) The bootstrap process $(X_t^*)_{t\in\mathbb{Z}}$ is strictly stationary with probability tending to one and takes its values in $\mathbb{R}^d$. Additionally,

$$P^*\left(\sup_{\omega:\, \|\omega\|_2 \le K} \left| \frac{1}{n} \sum_{t=1}^n e^{i\omega^T X_t^*} - E_{\theta_0} e^{i\omega^T X_t} \right| > \epsilon \right) \xrightarrow{P} 0 \qquad \forall K < \infty, \epsilon > 0, \tag{20}$$

that is, the empirical bootstrap measure converges weakly to P^{X_0} in probability.
(ii) The kernels of the bootstrap statistics $h^*\,:\ \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ are symmetric, nonnegative definite, and equicontinuous on compacta in probability, that is, $\forall\ K < \infty, \varepsilon > 0$, $\exists\ \delta > 0$ such that

$$P\left(\sup_{x_0,y_0:\,\|x_0\|_2,\|y_0\|_2\le K} \sup_{x,y:\,\|x-x_0\|_2,\|y-y_0\|_2\le\delta} |h^*(x,y)-h^*(x_0,y_0)|>\epsilon \right) \underset{n\to\infty}{\to} 0.$$

(iii) $h^*(x,y) \overset{P}{\to} h(x,y) \quad \forall\, x,y \in supp(P^{X_0})$.
(iv) $E^*h^*(X_0^*,X_0^*) \overset{P}{\to} Eh(X_0,X_0)$.
(v) $E^*(h^*(x,X_t^*)|X_{t-1}^*,\ldots,X_1^*)=0 \;\; a.s. \quad \forall\, x\in supp(P^{*X_0^*})$.

While part (ii) to (iv) of (B2) are specific conditions on the kernel h^* which have to be checked in a case by case manner, part (i) follows from (14); see also lemma 4.1 in Leucht and Neumann (2013). Theorem 4.1 in Leucht and Neumann (2013) provides a general consistency result. Actually, under conditions (B1) and (B2),

$$V_n^* \overset{d}{\to} Y \qquad \text{and} \qquad U_n^* \overset{d}{\to} Y - Eh(X_0,X_0) \quad \text{in probability.}$$

If additionally $P(h(X_0,X_0')\neq 0)>0$, for X_0' being an independent copy of X_0, then

$$\sup_{x\in\mathbb{R}}|P^*(U_n^*\le x)-P(U_n\le x)| \overset{P}{\to} 0 \quad \text{and} \quad \sup_{x\in\mathbb{R}}|P^*(V_n^*\le x)-P(V_n\le x)| \overset{P}{\to} 0.$$

5 | PROOFS

Proof of Lemma 1. A comparison of coefficients in (3) reveals that it suffices to find strictly positive constants $\gamma_1,\ldots,\gamma_p,\delta_1,\ldots,\delta_q$ and some $\kappa<1$ such that the following inequalities are satisfied:

$$\begin{aligned}
(\gamma_1+\delta_1)\ c_1+\gamma_2 &\le \kappa\ \gamma_1\\
&\vdots\\
(\gamma_1+\delta_1)\ c_{p-1}+\gamma_p &\le \kappa\ \gamma_{p-1}\\
(\gamma_1+\delta_1)\ c_p &\le \kappa\ \gamma_p\\
(\gamma_1+\delta_1)\ d_1+\delta_2 &\le \kappa\ \delta_1\\
&\vdots\\
(\gamma_1+\delta_1)\ d_{q-1}+\delta_q &\le \kappa\ \delta_{q-1}\\
(\gamma_1+\delta_1)\ d_q &\le \kappa\ \delta_q.
\end{aligned} \tag{21}$$

We set $\gamma_1+\delta_1=1$. Let $\epsilon=(1-L)/(p+q)$, where $L=\sum_{i=1}^p c_i+\sum_{j=1}^q d_j$. We consider the following system of equations.

$$\begin{aligned}
c_p+\epsilon &= \gamma_p\\
c_{p-1}+\gamma_p+\epsilon &= \gamma_{p-1}\\
&\vdots\\
c_1+\gamma_2+\epsilon &= \gamma_1\\
d_q+\epsilon &= \delta_q\\
d_{q-1}+\delta_q+\epsilon &= \delta_{q-1}\\
&\vdots\\
d_1+\delta_2+\epsilon &= \delta_1.
\end{aligned}$$

It is obvious that this system of equations has a unique solution with strictly positive $\gamma_1, \ldots, \gamma_p, \delta_1, \ldots, \delta_q$. Moreover, it follows from

$$\sum_{i=1}^{p} c_i + \sum_{j=1}^{q} d_j + \sum_{i=2}^{p} \gamma_i + \sum_{j=2}^{q} \delta_j + (p+q)\epsilon = \sum_{i=1}^{p} \gamma_i + \sum_{j=1}^{q} \delta_j,$$

and $\sum_{i=1}^{p} c_i + \sum_{j=1}^{q} d_j = L$ that $\gamma_1 + \delta_1 = 1$, as required. Therefore, we see that, with such a choice of $\gamma_1, \ldots, \gamma_p, \delta_1, \ldots, \delta_q$, the following strict inequalities are fulfilled.

$$\begin{aligned} c_1 + \gamma_2 &< \gamma_1 \\ &\vdots \\ c_{p-1} + \gamma_p &< \gamma_{p-1} \\ c_p &< \gamma_p \\ d_1 + \delta_2 &< \delta_1 \\ &\vdots \\ d_{q-1} + \delta_q &< \delta_{q-1} \\ d_q &< \delta_q. \end{aligned}$$

Choosing $\kappa = \max\{(c_1 + \gamma_2)/\gamma_1, \ldots, (c_{p-1} + \gamma_p)/\gamma_{p-1}, c_p/\gamma_p, (d_1 + \delta_2)/\delta_1, \ldots, (d_{q-1} + \delta_q)/\delta_{q-1}, d_q/\delta_q\}$ we obtain that the system of inequalities (21) is satisfied. ■

Proof of Proposition 2. Let Q and Q' be arbitrary probability measures supported in S and let ξ be the optimal coupling of Q and Q' w.r.t. the Kantorovich distance, that is,

$$\mathcal{K}(Q, Q') = \int_{S \times S} \Delta(z, z') \xi(dz, dz').$$

Then $\xi \tilde{\pi}_\theta^Z$ is a coupling of $Q\pi_\theta^Z$ and $Q'\pi_\theta^Z$ and it follows from Proposition 1(i) that

$$\begin{aligned} \mathcal{K}(Q\pi_\theta^Z, Q'\pi_\theta^Z) &\le \int \Delta(u, u') \xi \tilde{\pi}_\theta^Z(du, du') \\ &= \int \left[\int \Delta(u, u') \tilde{\pi}_\theta^Z((z, z'), (du, du')) \right] \xi(dz, dz') \\ &\le \kappa \int \Delta(z, z') \xi(dz, dz') = \kappa\ \mathcal{K}(Q, Q'). \end{aligned}$$

■

Proof of Corollary 1. Let

$$\mathcal{P} = \left\{ Q : \quad Q \text{ is a probability distribution based in } S, \int_S \sum_{i=1}^{p+q} |x_i|\ Q(dx) < \infty \right\}.$$

It is well-known that the space $\mathcal{P}$ equipped with the Kantorovich metric $\mathcal{K}$ is complete. Since by Proposition 2 the mapping π_θ^Z is contractive it follows by the Banach fixed point theorem that the Markov kernel π_θ^Z admits a unique fixed point Q, i.e. $Q\pi_\theta^Z = Q$. In other words, Q is the unique stationary probability distribution of the process $(Z_t)_{t \in \mathbb{Z}}$. ■

Proof of Lemma 2. (i) We prove this assertion by induction. We begin with $k=1$. Recall that it follows from (A1) that

$$\lambda_t \;\leq\; \sum_{i=1}^{p} c_i\, X_{t-i} + \sum_{j=1}^{q} d_j\, \lambda_{t-j} + C^{(0)}, \tag{22}$$

holds for all $t \in \mathbb{N}$. Let $M_t := \max\left\{\mathbb{E}X_t, \ldots, \mathbb{E}X_{t-p+1}, \mathbb{E}\lambda_t, \ldots, \mathbb{E}\lambda_{t-q+1}\right\}$. We obtain from (22) that

$$\mathbb{E}X_t = \mathbb{E}\lambda_t \;\leq\; L\, M_{t-1} + C^{(0)},$$

and therefore $M_t \leq \max\{M_{t-1}, LM_{t-1} + C^{(0)}\} \leq \max\{M_{t-1}, C^{(0)}/(1-L)\}$. Hence,

$$\max\left\{M_t, C^{(0)}/(1-L)\right\} \;\leq\; \max\left\{M_{t-1}, C^{(0)}/(1-L)\right\},$$

again for all $t \in \mathbb{N}$. This implies that

$$\mathbb{E}X_t = \mathbb{E}\lambda_t \;\leq\; C_1 = D_1 \;:= \max\left\{M_0, C^{(0)}/(1-L)\right\} \qquad \forall t \in \mathbb{N}.$$

Suppose now that, for all $t \in \mathbb{N}$ and $l=1, \ldots, k-1$,

$$\mathbb{E}X_t^l \;\leq\; C_l \qquad \text{and} \qquad \mathbb{E}\lambda_t^l \;\leq\; D_l$$

hold true, where $C_1, D_1, \ldots, C_{k-1}, D_{k-1}$ are finite constants which depend only on $c_1, \ldots, c_p, d_1, \ldots, d_q$ and $\max\{\mathbb{E}X_0^{k-1}, \ldots, \mathbb{E}X_{1-p}^{k-1}, \mathbb{E}\lambda_0^{k-1}, \ldots, \mathbb{E}\lambda_{1-q}^{k-1}\}$. We obtain from (22) that

$$\begin{aligned}
\mathbb{E}\lambda_t^k &\leq \mathbb{E}\left[\left(\sum_{i=1}^{p} c_i\, X_{t-i} + \sum_{j=1}^{q} d_j\, \lambda_{t-j} + C^{(0)}\right)^k\right] \\
&= \sum_{i_1+\ldots+i_p+j_1+\ldots+j_q=k} c_1^{i_1} \ldots c_p^{i_p}\, d_1^{j_1} \ldots d_q^{j_q}\, \mathbb{E}\left[X_{t-1}^{i_1} \ldots X_{t-p}^{i_p}\, \lambda_{t-1}^{j_1} \ldots \lambda_{t-q}^{j_q}\right] \\
&\quad + \sum_{l=1}^{k} \binom{k}{l} C^{(0)^l}\, \mathbb{E}\left[\left(\sum_{i=1}^{p} c_i\, X_{t-i} + \sum_{j=1}^{q} d_j\, X_{t-j}\right)^{k-l}\right].
\end{aligned} \tag{23}$$

Since, by assumption, all moments up to order $k-1$ are bounded we obtain that the second term on the right-hand side of (23) can be bounded by some constant which only depends on $c_1, \ldots, c_p, d_1, \ldots, d_q$ and $\max\{\mathbb{E}X_0^k, \ldots, \mathbb{E}X_{1-p}^k, \mathbb{E}\lambda_0^k, \ldots, \mathbb{E}\lambda_{1-q}^k\}$. To estimate the first term on the right-hand side of (23), note that, for $X \sim \text{Pois}(\lambda)$,

$$EX^k = \sum_{i=1}^{k} \lambda^i\, S(k,i), \tag{24}$$

where $S(k,i)$ denotes a Stirling number of the second kind; see for example, Riordan (1937, Equation 3.4). $S(k,i)$ is the number of ways to partition a set of k objects into i nonempty subsets. In what follows, it is only relevant that $S(k,k)=1$ for all $k \in \mathbb{N}$. We obtain that

$$\mathbb{E}\left[X_{t-1}^{i_1} \, \dots \, X_{t-p}^{i_p} \, \lambda_{t-1}^{j_1} \, \dots \, \lambda_{t-q}^{j_q}\right] \leq \left(\mathbb{E}X_{t-1}^k\right)^{i_1/k} \dots \left(\mathbb{E}X_{t-p}^k\right)^{i_p/k} \left(\mathbb{E}\lambda_{t-1}^k\right)^{j_1/k} \dots \left(\mathbb{E}\lambda_{t-q}^k\right)^{j_q/k}$$
$$\leq \max\left\{\mathbb{E}\lambda_{t-1}^k, \, \dots \, , \mathbb{E}\lambda_{t-(p\vee q)}^k\right\} + C^{(k)} \qquad \forall t > p,$$

where $C^{(k)}$ depends only on moments of order up to $k-1$. Therefore, we obtain from (23) the recursion

$$\mathbb{E}\lambda_t^k \;\leq\; L^k \; \max\{\mathbb{E}\lambda_{t-1}^k, \, \dots \, , \mathbb{E}\lambda_{t-(p\vee q)}^k\} + \tilde{C}^{(k)} \qquad \forall t > p,$$

where $\tilde{C}^{(k)} < \infty$ depends only on $c_1, \, \dots \, , c_p, d_1, \, \dots \, , d_q$ and $\max\{\mathbb{E}X_0^k, \, \dots \, , \mathbb{E}X_{1-p}^k, \mathbb{E}\lambda_0^k, \, \dots \, , \mathbb{E}\lambda_{1-q}^k\}$ Let $M_t^{(k)} := \max\left\{\mathbb{E}\lambda_t^k, \, \dots \, , \mathbb{E}\lambda_{t-(p\vee q)+1}^k\right\}$. Then $M_t^{(k)} \leq \max\{M_{t-1}^{(k)}, L^k M_{t-1}^{(k)} + \tilde{C}^{(k)}\} \leq \max\{M_{t-1}^{(k)}, \tilde{C}^{(k)}/(1-L^k)\}$. Hence,

$$\max\left\{M_t^{(k)}, \tilde{C}^{(k)}\right\} \;\leq\; \max\left\{M_{t-1}^{(k)}, \tilde{C}^{(k)}\right\} \qquad \forall t > p,$$

which implies that there exists some $D_k < \infty$ such that

$$\mathbb{E}\lambda_t^k \;\leq\; D_k \qquad \forall t \in \mathbb{N}.$$

Using (24) once more we obtain that there exists some $C_k < \infty$ such that

$$\mathbb{E}X_t^k \;\leq\; C_k \qquad \forall t \in \mathbb{N}.$$

(ii) Let $(Z_t)_{t\in\mathbb{N}_0}$ be a process with pre-sample values $Z_0 = (X_0, \, \dots \, , X_{1-p}, \lambda_0, \, \dots \, , \lambda_{1-q})$ as above and let $(Z_t^{(0)})_{t\in\mathbb{Z}}$ be a stationary version of the process. Since the law of Z_t converges to its stationary limit we conclude that $\lambda_t \stackrel{d}{\rightarrow} \lambda_t^{(0)}$ and $X_t \stackrel{d}{\rightarrow} X_t^{(0)}$. By theorem III.6.31 in Pollard (1984, p. 58) we can find a coupling of these random variables such that $\lambda_t \stackrel{a.s.}{\rightarrow} \lambda_t^{(0)}$ and $X_t \stackrel{a.s.}{\rightarrow} X_t^{(0)}$. Hence, we conclude by Fatou's lemma that

$$\mathbb{E}(\lambda_t^{(0)})^k \;\leq\; \liminf_{t\to\infty} \mathbb{E}\lambda_t^k,$$

and

$$\mathbb{E}(X_t^{(0)})^k \;\leq\; \liminf_{t\to\infty} \mathbb{E}X_t^k,$$

which proves the second assertion. ∎

Proof of Theorem 1. It follows from (4) to (7) that

$$\beta^X(k,n) \leq \sum_{l=0}^{\infty} \tilde{P}\left(\tilde{X}_{n+k+l} \neq \tilde{X}'_{n+k+l}\right)$$
$$\leq \frac{1}{\gamma_1} \sum_{l=0}^{\infty} \tilde{E}\Delta(\tilde{Z}_{n+k+l}, \tilde{Z}'_{n+k+l})$$
$$\leq \frac{1}{\gamma_1} \, \frac{\kappa^n}{1-\kappa} \, \tilde{E}\Delta(\tilde{Z}_k, \tilde{Z}'_k).$$

(i) and (ii) of Lemma 2 show that $\sup_k \left\{ \tilde{E}\Delta(\tilde{Z}_k, \tilde{Z}_k') \right\}$ is bounded, which completes the proof. ■

Proof of Theorem 2. This proof is analogous to that of Theorem 1 and therefore omitted. ■

Proof of Theorem 3. The arguments are already given in Section 3.3. ■

ACKNOWLEDGEMENTS

I thank two anonymous referees and an Associate Editor for helpful comments on an earlier version of this paper.

ORCID

Michael H. Neumann https://orcid.org/0000-0002-5783-831X

REFERENCES

Aknouche, A., & Francq, C. (2018). Count and duration time series with equal conditional stochastic and mean orders. Forthcoming in *Econometric Theory*, Working paper Retrieved from https://mpra.ub.uni-muenchen.de/97392/

Bickel, P. J., & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, *9*(6), 1196–1217.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Davis, R. A., Holan, S. H., Lund, R., & Ravishanker, N. (Eds.). (2016). *Handbooks of modern statistical methods. Handbook of Discrete-Valued Time series*. London, UK: Chapman & Hall/CRC Press.

Dehling, H., & Mikosch, T. (1994). Random quadratic forms and the bootstrap for *U*-statistics. *Journal of Multivariate Analysis*, *51*, 392–413.

Doukhan, P., Leucht, A., & Neumann, M. H. (2020). *Mixing properties of non-stationary INGARCH(1,1) processes*. Working paper Retrieved from https://arxiv.org/abs/2011.05854

Doukhan, P., & Neumann, M. H. (2019). Absolute regularity of semi-contractive GARCH-type processes. *Journal of Applied Probability*, *56*, 91–115.

Eberle, A. (2020). *Markov processes, Lecture Notes*. Germany: University of Bonn. https://uni-bonn.sciebo.de/s/kzTUFff5FrWGAay#pdfviewer.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Fokianos, K. (2012). *Count time series*. In: T. Subba Rao, S. Subba Rao, and C. R. Rao. *Time series: Methods and applications, handbook of statistics 30*, Elsevier, Amsterdam, Netherlands: pp. 315–347.

Fokianos, K., & Neumann, M. H. (2013). A goodness-of-fit test for Poisson count processes. *Electronic Journal of Statistics*, *7*, 793–819.

Fokianos, K., & Tjøstheim, D. (2011). Log-linear Poisson autoregression. *Journal of Multivariate Analysis*, *102*, 563–578.

Fokianos, K., & Tjøstheim, D. (2012). Nonlinear Poisson autoregression. *Annals of the Institute for Statistical Mathematics*, *64*, 1205–1225.

Francq, C., & Zakoïan, J.-M. (2010). *GARCH models: Structure, statistical inference and financial applications*. Chichester, West Sussex: Wiley.

Freedman, D. A. (1981). Bootstrapping regression models. *Annals of Statistics*, *9*(6), 1218–1228.

Jentsch, C., & Weiss, C. H. (2019). Bootstrapping INAR models. *Bernoulli*, *25*, 2359–2408.

Leucht, A., & Neumann, M. H. (2009). Consistency of general bootstrap methods for degenerate U-and V-type statistics. *Journal of Multivariate Analysis*, *100*, 1622–1633.

Leucht, A., & Neumann, M. H. (2013). Dependent wild bootstrap for degenerate U- and V-statistics. *Journal of Multivariate Analysis*, *117*, 257–280.

Leucht, A., Neumann, M. H., & Kreiss, J.-P. (2015). A model specification test for GARCH(1,1) processes. *Scandinavian Journal of Statistics*, *42*, 1167–1193.

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

Pollard, D. (1984). *Convergence of stochastic processes*. New York, NY: Springer.

Riordan, J. (1937). Moment recurrence relations for binomial, Poisson and hypergeometric frequency distributions. *The Annals of Mathematical Statistics*, *8*(2), 103–111.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge, MA: Cambridge University Press.

How to cite this article: Neumann MH.. Bootstrap for integer-valued GARCH(p, q) processes. *Statistica Neerlandica*. 2021;1–21. https://doi.org/10.1111/stan.12238