Bernoulli **25**(3), 2019, 2359–2408
https://doi.org/10.3150/18-BEJ1057

Bootstrapping INAR models

CARSTEN JENTSCH[1] and CHRISTIAN H. WEISS[2]

[1]*Technische Universität Dortmund, Fakultät Statistik, D-44221 Dortmund, Germany.*
E-mail: jentsch@statistik.tu-dortmund.de
[2]*Helmut Schmidt University, Department of Mathematics and Statistics, PO Box 700822, D-22008 Hamburg, Germany. E-mail: weissc@hsu-hh.de*

Integer-valued autoregressive (INAR) models form a very useful class of processes to deal with time series of counts. Statistical inference in these models is commonly based on asymptotic theory, which is available only under additional parametric conditions and further restrictions on the model order. For general INAR models, such results are not available and might be cumbersome to derive. Hence, we investigate how the INAR model structure and, in particular, its similarity to classical autoregressive (AR) processes can be exploited to develop an asymptotically valid bootstrap procedure for INAR models. Although, in a common formulation, INAR models share the autocorrelation structure with AR models, it turns out that (a) consistent estimation of the INAR coefficients is not sufficient to compute proper 'INAR residuals' to formulate a valid model-based bootstrap scheme, and (b) a naïve application of an AR bootstrap will generally fail. Instead, we propose a general INAR-type bootstrap procedure and discuss parametric as well as semi-parametric implementations. The latter approach is based on a joint semi-parametric estimator of the INAR coefficients and the innovations' distribution. Under mild regularity conditions, we prove bootstrap consistency of our procedure for statistics belonging to the class of functions of generalized means. In an extensive simulation study, we provide numerical evidence of our theoretical findings and illustrate the superiority of the proposed INAR bootstrap over some obvious competitors. We illustrate our method by an application to a real data set about iceberg orders for the Lufthansa stock.

Keywords: bootstrap consistency; functions of generalized means; INAR residuals; parametric bootstrap; semi-parametric bootstrap; semi-parametric estimation; time series of counts

1. Introduction and motivation

In recent years, the interest in time series of counts has increased steadily. Among the models proposed for dependent count data, integer-valued autoregressive (INAR) time series processes enjoy great popularity and have been used extensively in the statistical literature. As introduced by McKenzie [37] and Al-Osh and Alzaid [1], (strictly) stationary INAR processes of first-order (INAR(1)) $(X_t, t \in \mathbb{Z})$ are defined to follow the recursion

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \qquad t \in \mathbb{Z}, \tag{1.1}$$

where $\alpha \in (0, 1)$ and $(\epsilon_t, t \in \mathbb{Z})$ is an independent and identically distributed (i.i.d.) process with $\epsilon_t \sim G$, where G is a distribution with range $\mathbb{N}_0 = \{0, 1, \ldots\}$ such that $\mu_\epsilon = E(\epsilon_t) < \infty$. As coined by Steutel and van Harn [49], the operator "$\circ$" in (1.1) is called *binomial thinning*

1350-7265

operator. For a random variable X_{t-1} having range $\mathbb{N}_0$, it is defined by

$$\alpha \circ X_{t-1} = \sum_{j=1}^{X_{t-1}} Z_j^{(t)}, \tag{1.2}$$

where $(Z_j^{(t)}, j \in \mathbb{N}, t \in \mathbb{Z})$ are i.i.d. binary (Bernoulli-distributed) random variables independent of $(\epsilon_t, t \in \mathbb{Z})$ with $P(Z_j^{(t)} = 1) = \alpha$. That is, $Z_j^{(t)} \sim \text{Bin}(1, \alpha)$ and, conditionally on X_{t-1}, we have $\alpha \circ X_{t-1} \sim \text{Bin}(X_{t-1}, \alpha)$, where $\text{Bin}(n, \pi)$ denotes the binomial distribution with parameters n and π. The binomial thinning operation in (1.2) allows to interpret (1.1) as a branching process with immigration: with X_{t-1} costumers waiting for service at time $t-1$, each of them is served with probability $1-\alpha$ during the period $(t-1, t]$ such that $\alpha \circ X_{t-1}$ customers remain in the queue ('survivors') and ϵ_t new customers arrive ('immigrants').

In the literature, the definition of INAR models of first order in (1.1) has been extended to general lag order $p \in \mathbb{N}$ essentially in two ways. Alzaid and Al-Osh [2] proposed INAR(p) processes that maintain the nice interpretation of customers waiting for service of the INAR(1) model in (1.1). However, the formulation of INAR(p) processes proposed by Du and Li [14] has been preferred and followed by most researchers; see, for example, Latour [36], Silva and Silva [48] or Drost, van den Akker and Werker [13]. In their formulation, which we will use throughout this paper, INAR(p) processes $(X_t, t \in \mathbb{Z})$ are defined to follow the recursion

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \cdots + \alpha_p \circ X_{t-p} + \epsilon_t, \qquad t \in \mathbb{Z}, \tag{1.3}$$

where $(\epsilon_t, t \in \mathbb{Z})$ as before and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)' \in (0,1)^p$ such that $\sum_{i=1}^p \alpha_i < 1$ and

$$\alpha_i \circ X_{t-i} = \sum_{j=1}^{X_{t-i}} Z_j^{(t,i)}, \tag{1.4}$$

where $(Z_j^{(t,i)}, j \in \mathbb{N}, t \in \mathbb{Z})$, $i = 1, \dots, p$, are mutually independent i.i.d. binary random variables with $Z_j^{(t,i)} \sim \text{Bin}(1, \alpha_i)$ that are independent of $(\epsilon_t, t \in \mathbb{Z})$.

The main reason why INAR(p) processes as proposed by Du and Li [14] are usually preferred in the literature is as follows. If $\sigma_\epsilon^2 = \text{Var}(\epsilon_t) < \infty$ holds, their autocorrelation structure turns out to be exactly the same as that of an autoregressive process of order p (AR(p)) with corresponding coefficients, that is,

$$X_t = \nu + \alpha_1 \cdot X_{t-1} + \cdots + \alpha_p \cdot X_{t-p} + e_t, \qquad t \in \mathbb{Z}, \tag{1.5}$$

where $\nu \in \mathbb{R}$ is an intercept term and $(e_t, t \in \mathbb{Z})$ is a white noise process with $e_t \sim (0, \sigma_e^2)$. Note that we distinguish INAR innovations ϵ_t and AR errors e_t throughout the paper. In contrast, the autocorrelation structure of INAR(p) processes in the set-up of Alzaid and Al-Osh [2] corresponds to that of an autoregressive moving-average ARMA($p, p-1$) process; see also Latour [36].

Statistical inference for count time series and, in particular, within the framework of INAR models of the form (1.3) is a general task. This concerns the construction of reliable confidence intervals for certain statistics of interest, but also the application of diagnostic tools that allow the identification of appropriate models for a given count time series, and checking the adequacy of a fitted model; see Jung, McCabe and Tremayne [25], Weiß [52]. Such tools for INAR models are commonly based on asymptotic theory and usually rely on rather restrictive model assumptions. In particular, the most prominent case of a Poisson INAR(1) model has been addressed extensively in the literature. For example, the residual diagnostics proposed by Freeland and McCabe [18] or the tests for overdispersion and zero inflation by Weiß, Puig and Homburg [53] only apply to INAR processes of order $p = 1$ and with Poisson-distributed innovations, also see Appendix E.3 for such results. For general INAR models, asymptotic theory is usually not available and might be cumbersome to derive. Only few special cases as, for example, the limiting distribution of Yule–Walker estimators has been derived for the general case of model order p and without additional assumptions on the distribution, see Silva and Silva [48]. To be able to do statistical inference for a general class of statistics and diagnostic tools also beyond Poisson INAR(1), a general bootstrap scheme for INAR processes of any order and without the need for specifying the innovations' distribution will be very helpful. Furthermore, finite sample improvements might be observed when using suitable bootstrap techniques in cases where asymptotic results in principle already allow statistical inference. Also the distribution of parameter estimators has been studied in similarly restrictive contexts (Freeland and McCabe [17], Bu, McCabe and Hadri [6], Weiß and Schweer [54]) such that resulting approaches for bias correction, approximation of standard errors or confidence intervals are rather limited in scope. This generally leads to the question of how to do bootstrapping for INAR count time series?

As stationary INAR(p) models are always assured to be geometrically strong mixing (Doukhan, Fokianos and Li [11,12]) and they are just a parsimonious form of Markov processes with countable state space $\mathbb{N}_0$, two (groups of) non-parametric bootstrap procedures can be applied consistently to INAR time series data. These are block bootstraps and Markovian bootstraps. Block bootstraps are generally applicable under strong mixing (see, e.g., Künsch [35]), but they require the selection of a tuning parameter (block length) and tend to perform inferior in small samples. Markovian bootstraps suffer from a huge number of parameters (transition probabilities), especially for INAR processes with their infinite range and a possibly large Markov order. For prediction in INAR models, block bootstrap techniques have been used, for example, by Jung and Tremayne [26] to estimate the standard errors. Potential further issues of (block) bootstrapping for discrete-valued time series are discussed in Jentsch and Leucht [21].

In general, if a certain structure is available for a DGP (e.g., in form of a parametric model), it will be advisable to make use of such structural assumptions to construct a tailor-made bootstrap procedure. Usually, this leads to bootstraps that will be superior to other competitive bootstrap methods that are more widely applicable. As INAR models share the same autocorrelation function $\rho(h) = \mathrm{Corr}(X_{t+h}, X_t)$, $h \in \mathbb{Z}$, the INAR(p) model in (1.3) can be considered as a (non-negative) integer-valued counterpart to the continuous-valued AR(p) model in (1.5). Hence, a seemingly reasonable idea is trying to adapt the widely used AR bootstrap to the INAR case; see, for example, Kreiß and Paparoditis [33] for an overview of residual-based AR bootstrapping. Actually, the INAR coefficients $\alpha_1, \dots, \alpha_p$ and the intercept ν can be estimated by well-known techniques from classical time series analysis as, for example, Yule–Walker (YW), Least-Squares (LS) or Maximum-Likelihood (ML) estimators; see, for example, Du and Li [14], Silva and Silva [48], Bu, McCabe and Hadri [6]. However, both models differ in one crucial point. Contrary to AR models, which belong to the class of linear time series processes, INAR models are *non-linear*. This non-linearity is caused by the (random) thinning operations "$\circ$" in (1.3) in comparison to the simple (deterministic) multiplications "$\cdot$" in (1.5).

In particular, with data $X_1, \dots, X_n$ at hand, this additional random source introduced by the thinning operation in INAR models does not allow to identify the innovations ϵ_t even in the case when the model coefficients $\alpha_1, \dots, \alpha_p$ are known, also see the discussion in Appendix C. This is in contrast to AR models, where the error terms e_t, $t = p+1, \dots, n$, can be easily reconstructed using (1.5). Moreover, if $\alpha_1, \dots, \alpha_p$ are not known, it is still possible in AR models to consistently estimate the error terms e_t by residuals $\widehat{e}_t$ as $\alpha_1, \dots, \alpha_p$ and ν can be consistently estimated by some $\widehat{\alpha}_1, \dots, \widehat{\alpha}_p$ and $\widehat{\nu}$ (with $\sqrt{n}$-rate). More precisely, we have

$$\widehat{e}_t = X_t - \widehat{\nu} - \widehat{\alpha}_1 X_{t-1} - \dots - \widehat{\alpha}_p X_{t-p} \xrightarrow{P} X_t - \nu - \alpha_1 X_{t-1} - \dots - \alpha_p X_{t-p} = e_t, \qquad (1.6)$$

where $\xrightarrow{P}$ denotes convergence in probability. The latter observation allows to make meaningful inference about the distribution of the errors e_t as well. In particular, it allows the valid application of residual-based bootstrap schemes as the residuals carry automatically the true distribution of the error terms, that is, it is not necessary to estimate explicitly the distribution of the innovations or its features as, for example, its variance. We refer to Section 2 for details. So to adapt the AR bootstrap to the INAR case, it would be necessary to construct an appropriate set of residuals from INAR time series data. In Appendix C, we address this seemingly straightforward task in more detail. It turns out that, in comparison to the AR case, it is not possible to get proper residuals from observed INAR data that are "close to" the true innovations $\epsilon_2, \dots, \epsilon_n$ and carry the true innovations' distribution to an (asymptotically) sufficient extent. Alternatively, as INAR and AR models share the autocorrelation structure, it might be advisable to just ignore the discreteness of count data and to apply straight ahead an AR bootstrap to capture the dependence in INAR data. The generality of such an approach to lead to asymptotically valid bootstrap approximations will be discussed in detail in Section 2. It turns out that consistent application is restricted to certain special cases, and the standard AR bootstrap is not appropriate for general statistics when applied to INAR time series data.

The main contribution of this paper is to propose a general INAR-type bootstrap procedure in Section 3, where we prove consistency under suitable meta assumptions and discuss parametric and semi-parametric implementations. The latter will be semi-parametric in the sense that it mimics the parametric structure of the binomial thinning operations in an INAR recursion without assuming any parametric family of distributions for the innovation process $(\epsilon_t, t \in \mathbb{Z})$. In Section 4, we show the results of an extensive simulation study that compares several bootstrap approaches including, for example, block and Markovian bootstraps to construct confidence intervals. We find that our semi-parametric INAR bootstrap is robust and tends to perform superior to relevant competitors. In a real-data application, we analyze iceberg orders from Lufthansa stock to illustrate our method. Section 5 concludes, and all proofs, details about the simulation study as well as further background material are deferred to Appendices A–E.

2. Ignoring discreteness: Residual-based AR bootstrap

In the case where the DGP is assumed to be a stationary AR(p) process following a recursion (1.5) with i.i.d. white noise, it is usually advisable to use a model-based bootstrap to resample observations $X_1, \dots, X_n$. A corresponding residual-based AR bootstrap is well-suited for statistical inference and can be used to consistently estimate the distribution of statistics $T_n = T_n(X_1, \dots, X_n)$ belonging to a large class of statistics under mild conditions; compare Kreiß [30–32], Bühlmann [7,8], Kreiß, Paparoditis and Politis [34] and the overview in Kreiß and Paparoditis [33]. For example, it allows to construct confidence intervals in numerous situations that are of great interest in the statistical literature. The general idea of the residual-based AR bootstrap is to fit an AR(p) model (1.5) to the (centered) time series data and to compute AR residuals. These residuals are approximately i.i.d., also see (1.6), which allows the application of Efron's i.i.d. bootstrap (Efron [15]) to get (asymptotically valid) bootstrap error terms. Then, these bootstrap error terms are used to generate bootstrap data according the fitted AR(p) model.

Despite the fact that the observations and the innovations in INAR models are always non-negative and integer-valued, they are still autoregressive-type DGPs. Hence, and as they even share the autocorrelation structure with AR models, one may just ignore that the residuals obtained from fitting AR models to INAR data are neither integer-valued nor assured to be non-negative, and apply the AR bootstrap to INAR data just straight ahead. More precisely, with data $X_1, \dots, X_n$ at hand, the AR bootstrap is as follows:[1]

Residual-based AR bootstrap

Step 1. Compute the centered observations $Y_t = X_t - \overline{X}$, where $\overline{X} = \frac{1}{n}\sum_{t=1}^n X_t$.

Step 2. Fit an autoregressive process of order p, that is, $Y_t = \sum_{i=1}^p \alpha_i Y_{t-i} + e_t$, to the centered data $Y_1, \dots, Y_n$. This leads to estimated AR coefficients $\widehat{\alpha}_1, \dots, \widehat{\alpha}_p$, which can be obtained, for example, from Yule–Walker estimates.

Step 3. Compute the estimated residuals $\widehat{e}_t = Y_t - \sum_{i=1}^p \widehat{\alpha}_i Y_{t-i}$, $t = p+1, \dots, n$, and center them, that is, compute $\widetilde{e}_t = \widehat{e}_t - \frac{1}{n-p}\sum_{t=p+1}^n \widehat{e}_t$.

Step 4. Generate bootstrap observations $Y_1^+, \dots, Y_n^+$ according to

$$Y_t^+ = \sum_{i=1}^p \widehat{\alpha}_i Y_{t-i}^+ + e_t^+,$$

where (e_t^+) are randomly and uniformly drawn from the centered residuals $\{\widetilde{e}_{p+1}, \dots, \widetilde{e}_n\}$. Defining $X_t^+ = Y_t^+ + \overline{X}$ leads to the AR bootstrap sample $X_1^+, \dots, X_n^+$.

In particular, in view of the analogy of INAR and AR models with respect to their autocorrelation structure, a naïve application of an AR bootstrap to INAR data clearly has some potential to lead to asymptotically valid bootstrap approximations. In particular, whenever a statistic of interest does depend exclusively on the second-order dependence structure of the

[1]Note that this approach is asymptotically equivalent to the approach, where intercept and AR parameters are estimated jointly, where no centering in Step 1 has to be applied.

DGP, heuristically, the AR bootstrap might be sufficient. The latter holds true for the sample mean $\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$, which depends exclusively on autocovariances in the limit. Precisely, let $(X_t, t \in \mathbb{Z})$ be an INAR(p) process with $\mu_X = E(X_t)$, $\gamma_X(h) = \mathrm{Cov}(X_{t+h}, X_t)$, $h \in \mathbb{Z}$ and suppose $2+\delta$ moments exist. Then, as INAR processes are always geometrically strong mixing (Doukhan, Fokianos and Li [11,12]), the central limit theorem (CLT) from Ibragimov [19] gives

$$\sqrt{T}(\overline{X} - \mu_X) \xrightarrow{d} \mathcal{N}(0, V), \qquad V = \sum_{h=-\infty}^{\infty} \gamma_X(h), \tag{2.1}$$

where "$\xrightarrow{d}$" denotes convergence in distribution. Indeed, the following lemma shows that an AR bootstrap applied to INAR data is generally consistent for the sample mean although the discrete nature of the underlying process is completely ignored; see Appendix D for details including the proof.

Lemma 2.1 (AR bootstrap consistency for the sample mean). *Suppose* $(X_t, t \in \mathbb{Z})$ *is a stationary* INAR(p) *model as in* (1.3) *with* $E(\epsilon_t^4) < \infty$. *Then, we have* $n\,\mathrm{Var}^+(\overline{Y}^+) \to V$ *and*

$$\sqrt{n}\overline{Y}^+ = \sqrt{n}\big(\overline{X}^+ - \overline{X}\big) \xrightarrow{d} \mathcal{N}(0, V)$$

in probability, respectively.

The above result is closely related to those obtained in Kreiß, Paparoditis and Politis [34]. By introducing the concept of companion processes, they show in their Theorem 3.1 that an AR (sieve) bootstrap that captures asymptotically the complete autocovariance structure of an underlying process, will always be consistent for the sample mean (and other statistics that depend exclusively on the autocovariance structure of the process) under very mild conditions.

However, the situation becomes much different for statistics that depend also on other distributional features of the DGP beyond second-order structure. To illustrate this, we consider the sample variance $\widehat{\gamma}(0) = n^{-1}\sum_{t=1}^{n}(X_t - \overline{X})^2$ for Poisson INAR(1) data in the following example, which shows that an AR bootstrap will be invalid already in this simple case; see Appendix D for a sketch of the proof and a numerical example that quantifies the AR bootstrap failure.

Example 2.2 (AR bootstrap inconsistency for the sample variance). Suppose $(X_t, t \in \mathbb{Z})$ is a stationary Poisson INAR(1) process as in (1.1) with $\epsilon_t \sim \mathrm{Poi}(\lambda)$, $\lambda > 0$ and thinning parameter $\alpha \in (0,1)$. This leads to $\mu_X = E(X_t) = \lambda/(1-\alpha)$, and the true limiting variance of the sample variance $\widehat{\gamma}(0)$ is given by

$$n\,\mathrm{Var}\big(\widehat{\gamma}(0)\big) \to 2\mu_X^2 \frac{1+\alpha^2}{1-\alpha^2} + \mu_X \frac{1+\alpha}{1-\alpha}, \tag{2.2}$$

see Weiß and Schweer [54]. However, if an ordinary AR bootstrap (of order one) is applied to INAR(1) data, we get

$$n\,\mathrm{Var}^+\big(\widehat{\gamma}^+(0)\big) \to 2\mu_X^2 \frac{1+\alpha^2}{1-\alpha^2} + \mu_X \frac{1-4\alpha^2+6\alpha^3-3\alpha^4}{(1-\alpha^2)^2} \tag{2.3}$$

in probability. A comparison of (2.2) and (2.3) shows that their respective second summands differ, leading to the conclusion that the AR bootstrap is not asymptotically valid.

A similar discrepancy is also observed for the sample autocorrelation function of a Poisson INAR(1) process. While

$$\sqrt{n}\big(\widehat{\rho}(1)-\rho(1)\big) \xrightarrow{d} \mathcal{N}\left(0, 1-\alpha^2+\frac{\alpha(1-\alpha)}{\mu_X}\right)$$

holds in this case (Weiß and Schweer [54]), we get for the AR bootstrap counterpart that

$$\sqrt{n}\big(\widehat{\rho}^+(1)-\widehat{\rho}(1)\big) \xrightarrow{d} \mathcal{N}\big(0, 1-\alpha^2\big) \qquad \text{in probability,}$$

see, for example, Brockwell and Davis [5], Example 7.2.3. The additional term $\alpha(1-\alpha)/\mu_X$ may be large for small values of the observations' mean μ_X. The latter observation of AR bootstrap inconsistency for autocorrelations is in sharp contrast to linear time series, where an AR (sieve) bootstrap will be generally consistent for autocorrelations (due to Bartlett's formula) although it might not be consistent for autocovariances.

Furthermore, an AR bootstrap that just ignores the fact the INAR observations and innovations are non-negative and integer-valued, is particularly not suitable at all for statistics that rely on the discreteness of the data. For example, the zero-probability $p_0 = P(X_t = 0)$ can be easily estimated by the relative frequency of zeros in a sample $X_1, \ldots, X_n$, but this will be not possible from AR bootstrap data $X_1^+, \ldots, X_n^+$ that is no longer discrete with range $\mathbb{N}_0$; compare also the simulation results in Section 4.1.

As the application of an ordinary AR bootstrap to INAR data is in general not satisfactory due to the issues raised above, there is a great need for INAR-type bootstrap procedures that allow a consistent application to a broad class of statistics.

3. INAR bootstraps

As observed in the previous section, an ordinary AR bootstrap cannot be valid for INAR models in general. It is obviously crucial to replicate the randomness of the thinning operation in the INAR recursion *and* to use appropriate bootstrap innovations. Hence, an INAR-type bootstrap is required rather than an AR-type bootstrap. However, in view of Appendix C, it is not possible to construct INAR residuals that carry the true innovations' distribution to a sufficient extent. Consequently, a general proposal of a model-based bootstrap that exploits the autoregressive nature of INAR processes and, at the same time, can be valid for general statistics is not straightforward.

However, it is not essential to use explicitly residuals from a fitted model to get bootstrap innovations for an INAR recursion. Instead, suitable random variables have to be used whose marginal distribution shares asymptotically all relevant features with the true innovations' distribution. Depending on the statistic of interest and on the conditions imposed on the DGP, these bootstrap innovations can be obtained in different ways.

In the following, we present a general INAR-type bootstrap scheme in Section 3.1, that does not specify the method used to estimate the INAR coefficients $\alpha_1, \ldots, \alpha_p$ and the innovations' distribution G. The subsequent formulation of a general INAR bootstrap scheme allows a unifying asymptotic treatment of e.g. parametric and semi-parametric implementations. In Section 3.2, we prove bootstrap consistency under a suitable set of meta-assumptions on corresponding estimators $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ and $\widehat{G} = (\widehat{G}(k), k \in \mathbb{N}_0)$. These assumptions require consistent estimation of $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ towards $\alpha_1, \ldots, \alpha_p$, but only the convergence of moments of $\widehat{G}$ to those of G up to some finite order. Although it will be generally more constructive to assume the whole distribution $\widehat{G}$ to converge towards G, we will discuss that an INAR bootstrap can indeed be consistent without this requirement for the case of the sample mean. In Sections 3.3 and 3.4, we propose parametric and semi-parametric implementations of the INAR bootstrap procedure.

3.1. The INAR bootstrap scheme

With data $X_1, \ldots, X_n$ at hand, a general INAR bootstrap scheme is defined as follows:

Step 1. Fit an INAR(p) process $X_t = \sum_{k=1}^{p} \alpha_k \circ X_{t-k} + \epsilon_t$ to the data to get estimates $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ for the INAR coefficients and $\widehat{G} = (\widehat{G}(k), k \in \mathbb{N}_0)$ for the innovations' distribution.

Step 2. Generate bootstrap observations $X_1^*, \ldots, X_n^*$ according to

$$X_t^* = \sum_{k=1}^{p} \widehat{\alpha}_k \circ^* X_{t-k}^* + \epsilon_t^*, \tag{3.1}$$

where "$\circ^*$" denotes (mutually independent) bootstrap binomial thinning operations and (ϵ_t^*) are i.i.d. random variables following the distribution $\widehat{G}$.

3.2. Bootstrap consistency

In this section, we specify the class of functions of generalized means that we consider here. It is worth mentioning that the considered class is actually richer for discrete-valued data and allows a treatment of statistics that are not covered in the continuous setup. Then, we state suitable assumptions on the INAR process and meta-assumptions on the estimation scheme for $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ and the innovations' distribution $\widehat{G}$ used to generate bootstrap innovations. Together, these conditions allow to prove bootstrap consistency results.

A statistic $T_n = T_n(X_1, \ldots, X_n)$ belongs to the class of functions of generalized means if for functions $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$, T_n can be represented as

$$T_n = f\left(\frac{1}{n_m} \sum_{t=1}^{n_m} \boldsymbol{g}(X_t, \ldots, X_{t+m-1})\right), \tag{3.2}$$

where $m \in \mathbb{N}$ (fixed) and $n_m = n - m + 1$. In the following, the functions f and $\boldsymbol{g}$ are assumed to fulfill the following smoothness conditions:

Assumption 1. The function f in (3.2) has continuous partial derivatives for all $\boldsymbol{y}$ in a neighborhood of $\boldsymbol{\xi} := E(\boldsymbol{g}(X_t, \dots, X_{t+m-1}))$, and the differentials $\sum_{i=1}^{d} \partial f(\boldsymbol{x})/\partial x_i|_{\boldsymbol{x}=\boldsymbol{\xi}}$ do not vanish. The function $\boldsymbol{g}$ has partial derivatives of some order h ($h \geq 1$), which satisfy a Lipschitz condition, that is, for all $i = 1, \dots, d$ and for all $(h_1, \dots, h_m) \in \mathbb{N}_0^m$ with $\sum_{u=1}^{m} h_u = h$, the derivative

$$\frac{\partial^h g_i(x_1, \dots, x_m)}{\partial^{h_1} x_1 \cdots \partial^{h_m} x_m}$$

is Lipschitz.

The above class of statistics is quite rich and contains as special cases, e.g., versions of sample autocovariances, sample autocorrelations, sample partial autocorrelations, Yule–Walker estimators, and the sample mean. In the context of bootstrap methods, this class of statistics has been considered by Künsch [35], Bühlmann [8], Kreiß, Paparoditis and Politis [34], Jentsch and Politis [22], Meyer and Kreiß [41] in the context of continuous-valued univariate and multivariate time series and by Meyer, Jentsch and Kreiß [40] for random fields observed on the lattice $\mathbb{Z}^2$.

It is worth noting that Assumption 1 was originally designed for continuous-valued random variables. In particular, it does not allow for discontinuous functions $\boldsymbol{g}$ ruling out, e.g., indicator functions. When dealing with integer-valued INAR processes $(X_t, t \in \mathbb{Z})$, i.e., if $P(X_t \in \mathbb{N}_0) = 1$ holds, Assumption 1 can be easily weakened. Precisely, the smoothness condition imposed on the function $\boldsymbol{g}$ can be relaxed to read as follows: The function $\boldsymbol{g}$ is such that there exists a function $\widetilde{\boldsymbol{g}} : \mathbb{R}^m \to \mathbb{R}^d$ with $\boldsymbol{g}(\boldsymbol{x}) = \widetilde{\boldsymbol{g}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{N}_0^m$, and $\widetilde{\boldsymbol{g}}$ fulfills the smoothness condition in Assumption 1. The latter condition covers, e.g., indicator functions (with $h = 1$) such that estimators based on relative frequencies can also be treated within our framework of Assumption 1. For example, the probability to observe a zero, that is, $p_0 := P(X_t = 0)$, which is a feature of the marginal distribution of the INAR process $(X_t, t \in \mathbb{Z})$, can be estimated from a sample $X_1, \dots, X_n$ by its relative frequency

$$\widehat{p}_0 := \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}(X_t = 0), \tag{3.3}$$

where $\mathbb{1}(A)$ denotes the indicator function that equals 1 if the condition A holds and 0 otherwise. Similarly, any unconditional, conditional, marginal or joint estimator of probabilities corresponding to the distribution of $X_1, \dots, X_m$ based on relative frequencies will be covered. However, other statistics as for example, sample quantiles are not included in this generalized class of functions; see also Jentsch and Leucht [21].

Example 3.1. The class of functions of generalized means with the relaxation of Assumption 1 discussed above includes, for instance,

(i) the relative frequency $\widehat{p}_0$ in (3.3) to estimate the probability to observe a zero in an INAR sample $p_0 = P(X_t = 0)$;

(ii) the conditional probability that i is observed one time step after j, that is, $p_{i|j} := P(X_{t+1} = i | X_t = j)$. Its canonical estimator is

$$\widehat{p}_{i|j} := \frac{\sum_{t=1}^{n-1} \mathbb{1}((X_{t+1}, X_t) = (i, j))}{\sum_{t=1}^{n-1} \mathbb{1}(X_t = j)}.$$

For the DGP of the INAR process, we make the following assumption:

Assumption 2. Let $(X_t, t \in \mathbb{Z})$ be a stationary INAR(p) process as in (1.3) with coefficients $\alpha_1, \dots, \alpha_p$ such that $\alpha_i \in (0, 1)$, $i = 1, \dots, p$ and $\sum_{i=1}^{p} \alpha_i < 1$. Further, the innovation process $(\epsilon_t, t \in \mathbb{Z})$ is i.i.d. with a distribution $G = \mathcal{L}(\epsilon_t)$ having range $\mathbb{N}_0$ such that $E_G(\epsilon_t^s) < \infty$ for some $s \in \mathbb{N}$ and $0 < G(0) < 1$ holds, where $G(k) = P(\epsilon_t = k)$, $k \in \mathbb{N}_0$.

These conditions on the underlying INAR process have been used in Drost, van den Akker and Werker [13]; see Section 3.4 for details. The condition $\alpha_i \in (0, 1)$, $i = 1, \dots, p$ with $\sum_{i=1}^{p} \alpha_i < 1$ and $E_G(\epsilon_t^s) < \infty$ for $s \geq 1$ is sufficient for the INAR recursion to have a strictly stationary solution. The assumption $0 < G(0) < 1$ ensures the possibility of X becoming 0 and not being always equal to 0, which is reasonable for virtually all applications. For deriving asymptotic theory in the following, we have to match the number s of existing moments of the innovations in Assumption 2 (and in Assumption 3 below) to h in Assumption 1 determining the smoothness of $\boldsymbol{g}$.

The INAR bootstrap scheme in Section 3.1 does not specify the estimation methods for the INAR coefficients and the innovations' distribution used to generate bootstrap data. Naturally, we have to impose certain additional meta-assumptions that eventually lead to valid bootstrap approximations. More precisely, we impose the following conditions on $\widehat{\alpha}_1, \dots, \widehat{\alpha}_p$ and $\widehat{G}$.

Assumption 3. Additional requirements concerning the INAR bootstrap scheme in Section 3.1:

(i) the estimators $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_p)'$ for the INAR coefficients in Step 1 satisfy

$$\sqrt{n}(\widehat{\alpha}_i - \alpha_i) = O_P(1), \qquad i = 1, \dots, p;$$

(ii) the bootstrap innovations' distribution $\widehat{G}$ used in Step 2 satisfies

$$E^*_{\widehat{G}}\left(\epsilon_t^{*l}\right) \xrightarrow{P} E_G\left(\epsilon_t^l\right) < \infty$$

for $l = 1, \dots, s$ with some $s \in \mathbb{N}$.

Under these assumptions, we are now able to prove bootstrap consistency in the following theorem, see Appendix A.1 for the proof.

Theorem 3.2 (INAR bootstrap consistency). *Suppose the statistic of interest $T_n = T_n(X_1, \dots, X_n)$ and the underlying* INAR(p) *process $(X_t, t \in \mathbb{Z})$ fulfill Assumptions 1 and 2 for some $h \in \mathbb{N}$*

and with $s = 2(h+2)$, respectively. Furthermore, if the estimators $\widehat{\alpha}_1, \dots, \widehat{\alpha}_p$ and the innovations' distribution $\widehat{G}$ used to generate bootstrap data $X_1^*, \dots, X_n^*$ satisfy Assumption 3 with $s = 2(h+2)$, we have

$$d_K\big(\mathcal{L}^*\big(\sqrt{n_m}\big(T_n^* - f\big(\boldsymbol{\xi}^*\big)\big)\big), \mathcal{L}\big(\sqrt{n_m}\big(T_n - f(\boldsymbol{\xi})\big)\big)\big) = o_P(1) \tag{3.4}$$

as $n \to \infty$, where $T_n^* = T_n(X_1^*, \dots, X_n^*)$ and $\boldsymbol{\xi}^* = E^*(\boldsymbol{g}(X_t^*, \dots, X_{t+m-1}^*))$.

Note that Assumption 3(ii) is to some extent tailor-made for the statistic of interest T_n and does not explicitly require that $\widehat{G}$ mimics asymptotically all properties of the true innovations' distribution G. Instead, bootstrap consistency might be achieved if only all asymptotically relevant features are mimicked. We illustrate this for the sample mean in the following example.

Example 3.3 (Sample mean). If $T_n(X_1, \dots, X_n) = \frac{1}{n}\sum_{t=1}^n X_t$ is the sample mean, it is sufficient for Assumption 3 to use any $\widehat{G}$ such that $E^*_{\widehat{G}}(\epsilon_t^*) \to E(\epsilon_t) = \mu_\epsilon$ and $\mathrm{Var}^*_{\widehat{G}}(\epsilon_t^*) \to \mathrm{Var}(\epsilon_t) = \sigma_\epsilon^2$ for the INAR bootstrap to be consistent. This is because, besides the INAR coefficients, only these first two moments appear in the limiting variance of $\overline{X}$. Hence, any distribution that satisfies this property is possible here (e.g., an appropriately chosen three-point distribution, see De Schepper and Heijnen [10]), although the true innovations' distribution might be different.

Under Assumptions 2 and 3, in the limit as $n \to \infty$, the DGP of the bootstrap process $(X_t^*, t \in \mathbb{Z})$ from Section 3.1 will follow an INAR(p) recursion

$$\widetilde{X}_t = \alpha_1 \circ \widetilde{X}_{t-1} + \alpha_2 \circ \widetilde{X}_{t-2} + \cdots + \alpha_p \circ \widetilde{X}_{t-p} + \widetilde{\epsilon}_t, \qquad t \in \mathbb{Z}, \tag{3.5}$$

where the marginal distribution of the innovation process $(\widetilde{\epsilon}_t)$ shares the first s moments with the true distribution G. The latter process $(\widetilde{X}_t, t \in \mathbb{Z})$ is the companion process in the sense of Kreiß, Paparoditis and Politis [34] that corresponds to the bootstrap scheme in Section 3.1.

To implement the INAR bootstrap procedure from Section 3.1, it is necessary to specify the estimation $\widehat{\alpha}_1, \dots, \widehat{\alpha}_p$ of the INAR coefficients in Step 1 as well as the choice of the distribution $\widehat{G}$ in Step 2. Of course, there are several possible choices. In the following, we address a fully parametric approach in Section 3.3, and a semi-parametric approach in Section 3.4.

3.3. Parametric implementation

In this section, we consider a fully parametric approach to specify the estimation $\widehat{\alpha}_1, \dots, \widehat{\alpha}_p$ of the INAR coefficients and the choice of the bootstrap innovations' distribution $\widehat{G}$. In the context of AR models, the use of a parametric bootstrap for model diagnostics goes back to Tsay [50]. Similarly, parametric (Poisson) INAR bootstraps have been used (among others) in Cardinal, Roy and Lambert [9] for forecasting INAR(p) processes, in Jung and Tremayne [27] for checking model adequacy, in Pavlopoulos and Karlis [44] for prediction and testing purposes, and in Meintanis and Karlis [38] for goodness-of-fit testing, while Schweer [46] used a parametric bootstrap for goodness-of-fit testing in the more general class of CLAR(1) models.

As AR and INAR models share the autocorrelation structure, suitable (parametric) estimators $\widehat{\alpha}_1, \dots, \widehat{\alpha}_p$ for the INAR coefficients in Step 1 of the bootstrap procedure in Section 3.1 are easily obtained, for example, by Yule–Walker estimation. Hence, it remains to estimate the innovations' distribution from the data to specify $\widehat{G} = \widehat{G}(X_1, \dots, X_n)$. By imposing a parametric family of distributions depending on a (finite-dimensional) parameter vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^l$, say, this parametric structure can be used to estimate the true innovations' distribution G from the INAR data. For a suitable parametric choice of $\widehat{G}$ in Step 1 of the bootstrap procedure, we make use of the following additional assumption on the innovations' distribution G.

Assumption 4. The marginal distribution $G = G_{\boldsymbol{\theta}_0}$ of the i.i.d. innovation process $(\epsilon_t, t \in \mathbb{Z})$ is a member of a parametric family $\{G_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta\}$ of distributions satisfying Assumption 2, where Θ is an open subset of $\mathbb{R}^l$ for some $l \in \mathbb{N}$. Suppose the map $\boldsymbol{\theta} \mapsto G_\theta$ is continuous on Θ, that is, for all $\boldsymbol{\theta} \in \Theta$ and all $\varepsilon > 0$, there exist a $\delta > 0$ such that for all $\boldsymbol{\theta}' \in \Theta$ with $|\boldsymbol{\theta} - \boldsymbol{\theta}'| < \delta$ we have $d_1(G_{\boldsymbol{\theta}}, G_{\boldsymbol{\theta}'}) := \sum_{k=0}^{\infty} |G_{\boldsymbol{\theta}}(k) - G_{\boldsymbol{\theta}'}(k)| < \varepsilon$. Further, suppose there exists a neighborhood $B_{\boldsymbol{\theta}_0,\delta} = \{\boldsymbol{\theta} \mid |\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta\} \subset \Theta$ of $\boldsymbol{\theta}_0$ such that $\sum_{k=0}^{\infty} k^s G_{\boldsymbol{\theta}}(k) < \infty$ holds uniformly on $B_{\boldsymbol{\theta}_0,\delta}$, and there exists an estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(X_1, \dots, X_n)$ for $\boldsymbol{\theta}_0$ such that $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1)$.

Under the parametric Assumption 4 on the innovations' distribution G, bootstrap innovations $\epsilon_1^*, \dots, \epsilon_n^*$ can be easily generated from $G_{\widehat{\boldsymbol{\theta}}}$ leading to the following bootstrap scheme and the subsequent direct corollary of Theorem 3.2.

Parametric INAR bootstrap

Step 1. Fit an INAR(p) process $X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \epsilon_t$ with $\epsilon_t \sim G_{\boldsymbol{\theta}}$ to the data to get (e.g., Yule–Walker) estimates $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_p)'$ and $\widehat{\boldsymbol{\theta}}$.

Step 2. Generate bootstrap observations $X_1^*, \dots, X_n^*$ according to

$$X_t^* = \sum_{i=1}^{p} \widehat{\alpha}_i \circ^* X_{t-i}^* + \epsilon_t^*,$$

where "$\circ^*$" denotes (mutually independent) bootstrap binomial thinning operations, and where (ϵ_t^*) are i.i.d. random variables following distribution $G_{\widehat{\boldsymbol{\theta}}}$, that is, $\widehat{G} := G_{\widehat{\boldsymbol{\theta}}}$ in Section 3.1.

Corollary 3.4 (Parametric INAR bootstrap consistency). *Suppose the statistic of interest $T_n = T_n(X_1, \dots, X_n)$ and the underlying* INAR(p) *process $(X_t, t \in \mathbb{Z})$ fulfill Assumption 1 for some $h \in \mathbb{N}$ and Assumption 2 with $s = 2(h+2)$, respectively. Furthermore, if the estimator $\widehat{\boldsymbol{\alpha}}$ satisfies Assumption 3(i), if the true innovations' distribution G fulfills Assumption 4 with $s = 2(h+2)$, and if $\widehat{G} = G_{\widehat{\boldsymbol{\theta}}}$ is used to generate the bootstrap data $X_1^*, \dots, X_n^*$, then we have*

$$d_K\big(\mathcal{L}^*\big(\sqrt{n_m}\big(T_n^* - f\big(\boldsymbol{\xi}^*\big)\big)\big), \mathcal{L}\big(\sqrt{n_m}\big(T_n - f(\boldsymbol{\xi})\big)\big)\big) = o_P(1) \tag{3.6}$$

as $n \to \infty$, where $T_n^* = T_n(X_1^*, \dots, X_n^*)$ and $\boldsymbol{\xi}^* = E^*(\boldsymbol{g}(X_t^*, \dots, X_{t+m-1}^*))$.

The latter result (its proof is provided by Appendix A.2) shows that a parametric INAR bootstrap is asymptotically valid for functions of generalized means if the parametric assumptions imposed on the innovation process $(\epsilon_t, t \in \mathbb{Z})$ are indeed fulfilled and if a suitable estimator $G_{\widehat{\theta}}$ for the innovations' distribution is used.

To investigate the potential of parametric INAR bootstraps to lead to valid approximations even if the innovations' distribution is misspecified, we consider the important special case of a Poisson INAR(1) model in more detail. It is very popular in the literature to assume a Poisson distribution for the innovation process, that is, $\mathcal{L}(\epsilon_t) = \mathrm{Poi}(\lambda)$, $\lambda > 0$. This is mainly because the Poisson distribution carries over to the marginal distribution of X_t if the INAR process $(X_t, t \in \mathbb{Z})$ is of order one. More precisely, this leads to the marginal distribution $\mathcal{L}(X_t) = \mathrm{Poi}(\lambda/(1-\alpha))$. Hence, in this Poisson case, the innovations' distribution $G_\lambda = \mathcal{L}(\epsilon_t)$ can be easily estimated by $G_{\widehat{\lambda}} = \mathrm{Poi}(\widehat{\lambda})$, where $\widehat{\lambda} = \overline{X}(1-\widehat{\alpha})$. This leads to the following parametric bootstrap scheme:

Example 3.5 (Poisson INAR(1) **Bootstrap).**

Step 1. Fit an INAR(1) process $X_t = \alpha \circ X_{t-1} + \epsilon_t$ with $\epsilon_t \sim \mathrm{Poi}(\lambda)$ to get estimates $\widehat{\alpha} = \widehat{\rho}(1)$ and $\widehat{\lambda} = \overline{X}(1-\widehat{\alpha})$.

Step 2. Generate bootstrap observations $X_1^*, \dots, X_n^*$ according to

$$X_t^* = \widehat{\alpha} \circ^* X_{t-1}^* + \epsilon_t^*,$$

where "$\circ^*$" denotes (mutually independent) bootstrap binomial thinning operations and (ϵ_t^*) are i.i.d. random variables following a Poisson distribution with parameter $\widehat{\lambda}$, that is, $\widehat{G} := \mathrm{Poi}(\widehat{\lambda})$ in Section 3.1.

Clearly, as $\sqrt{n}(\widehat{\lambda} - \lambda) = O_P(1)$ and as $E^*_{G_{\widehat{\lambda}}}(\epsilon_t^{*s}) \overset{P}{\to} E_{G_\lambda}(\epsilon_t^s) < \infty$ for all $s \in \mathbb{N}_0$, such a Poisson INAR(1) bootstrap will be generally consistent for functions of generalized means whenever the underlying process is indeed Poisson INAR(1). Otherwise, the bootstrap will usually fail. Hence, it is indeed restrictive to impose a parametric structure on the innovations' distribution. However, as will be illustrated in the following example, it is indeed not always necessary to specify the parametric family of innovations' distributions correctly to achieve bootstrap consistency.

Example 3.6 (Poisson INAR(1) **bootstrap: Sample mean).** In analogy to Lemma 2.1, for an INAR bootstrap to be consistent for the sample mean, it is sufficient to mimic the first two moments of the innovations' distribution; compare Example 3.3 and equation (2.1). Furthermore, Poisson distributions are equidispersed, that is, they have a variance equal to the mean. Hence, a Poisson INAR(1) bootstrap is consistent for the sample mean $\overline{X}$ if and only if the true innovations' distribution of an underlying INAR(1) process is equidispersed.

In many real-data applications, a Poisson distribution turns out to be not suitable as the data show overdispersion, that is, $\mathrm{Var}(X_t) > E(X_t)$, or the zero probability $p_0 = P(X_t = 0)$ is larger than for Poisson distributions, see Jazi, Jones and Lai [20], Schweer and Weiß [47], Weiß, Puig and Homburg [53]. Hence, imposing a Poisson distribution will often be too restrictive. In Section 4, Monte Carlo simulations illustrate that this can lead, for example, to considerable under-coverage of bootstrap confidence intervals. Hence, imposing parametric structure on the innovation process turns out to be a strong assumption that might not be appropriate in practice.

3.4. Semi-parametric implementation

A parametric implementation of the INAR bootstrap will generally be valid for functions of generalized means exclusively if the parametric class of innovations' distributions is correctly specified. To construct an INAR-type bootstrap procedure that does not rely on any parametric assumptions imposed on the innovations and is valid for general statistics, it is inevitable to estimate consistently the true innovations' distribution to mimic essentially all its features. Hence, a (non-parametric) estimation procedure is required to define a semi-parametric INAR bootstrap scheme. The approach will be semi-parametric in the sense that we stick to the parametric binomial thinning operation, but allow for non-parametric estimation of the innovations' distribution.

In a remarkable paper by Drost, van den Akker and Werker [13], for INAR(p) models in (1.3), the authors consider the semi-parametric joint estimation of the INAR coefficients $\alpha_1,\dots,\alpha_p$ *and* the innovations' distribution $G=\mathcal{L}(\epsilon_t)$. Their semi-parametric maximum likelihood estimator (SPMLE)

$$(\widehat{\boldsymbol{\alpha}}_{\mathrm{sp}},\widehat{G}_{\mathrm{sp}}):=\big(\widehat{\alpha}_{\mathrm{sp},1},\dots,\widehat{\alpha}_{\mathrm{sp},p},\widehat{G}_{\mathrm{sp}}(0),\widehat{G}_{\mathrm{sp}}(1),\widehat{G}_{\mathrm{sp}}(2),\dots\big)$$

is defined to maximize the conditional log-likelihood function (which uses the fact that the conditional distribution of X_t given $X_{t-1},\dots,X_{t-p}$ becomes a convolution of the binomial distributions $\mathrm{Bin}(X_{t-1},\alpha_1),\dots,\mathrm{Bin}(X_{t-p},\alpha_p)$ and of G), where the estimated probabilities $\widehat{G}_{\mathrm{sp}}(k)$ are equal to 0 for $k<\min\{X_t-\sum_{i=1}^{p}X_{t-i}\mid t=p+1,\dots,n\}$ and for $k>\max\{X_t\mid t=1,\dots,n\}$, also see Appendix E.1 for some more details. For the SPMLE, Drost, van den Akker and Werker [13] prove efficiency under some mild regularity conditions that cover rather general innovations' distributions with range $\mathbb{N}_0$. In particular, for the (infinite-dimensional) estimator $\widehat{G}_{\mathrm{sp}}=(\widehat{G}_{\mathrm{sp}}(k),k\in\mathbb{N}_0)$ of the innovations' distribution $G=(G(k),k\in\mathbb{N}_0)$, they prove ℓ_1-consistency and, moreover, weak convergence to a suitable Gaussian process. More precisely, in their Theorem 2, they prove

$$\sqrt{n}\big((\widehat{\boldsymbol{\alpha}}_{\mathrm{sp}},\widehat{G}_{\mathrm{sp}})-(\boldsymbol{\alpha},G)\big)\rightsquigarrow Z \tag{3.7}$$

under Assumption 2 plus a slightly stronger moment condition $E[\epsilon_t^{p+4}]<\infty$, where Z is a tight, Borel measurable, Gaussian process. The $(p+4)$th moment of G is needed to establish weak convergence of certain empirical processes. We refer to Drost, van den Akker and Werker [13] for details. Note that the sequence $\widehat{G}_{\mathrm{sp}}$ is always of bounded support for finite sample size n, that is, we have $\widehat{G}_{\mathrm{sp}}(k)=0$ whenever $k>\max\{X_1,\dots,X_n\}$. The weak convergence in (3.7) immediately implies

$$\begin{gathered}\sqrt{n}\big((\widehat{\alpha}_{\mathrm{sp},1},\dots,\widehat{\alpha}_{\mathrm{sp},p})-(\alpha_1,\dots,\alpha_p)\big)=O_P(1)\quad\text{and}\\ \sqrt{n}\sum_{k=0}^{\infty}\big|\widehat{G}_{\mathrm{sp}}(k)-G(k)\big|=O_P(1).\end{gathered} \tag{3.8}$$

This is the key result to establish bootstrap consistency for a semi-parametric INAR bootstrap that makes use of the estimated distribution $\widehat{G}_{\text{sp}}$ to generate bootstrap innovations. This leads to the following bootstrap scheme:

Semi-parametric INAR bootstrap

Step 1. Fit semi-parametrically an INAR(p) process $X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \epsilon_t$ to get estimated INAR coefficients $\widehat{\boldsymbol{\alpha}}_{\text{sp}} = (\widehat{\alpha}_{\text{sp},1}, \ldots, \widehat{\alpha}_{\text{sp},p})'$ and the estimator $\widehat{G}_{\text{sp}} = (\widehat{G}_{\text{sp}}(k), k \in \mathbb{N}_0)$.

Step 2. Generate bootstrap observations $X_1^*, \ldots, X_n^*$ according to

$$X_t^* = \widehat{\alpha}_{\text{sp},1} \circ^* X_{t-1}^* + \cdots + \widehat{\alpha}_{\text{sp},p} \circ^* X_{t-p}^* + \epsilon_t^*,$$

where "$\circ^*$" denotes (mutually independent) bootstrap binomial thinning operations and (ϵ_t^*) are i.i.d. random variables following $\widehat{G}_{\text{sp}}$.

Corollary 3.7 (Semi-parametric INAR bootstrap consistency). *Suppose the statistic of interest $T_n = T_n(X_1, \ldots, X_n)$ and the underlying* INAR(p) *process* $(X_t, t \in \mathbb{Z})$ *fulfill Assumption 1 for some $h \in \mathbb{N}$ and Assumption 2 with $s >$* $\max\{4(h+2), p+4\}$, *respectively. Then, if* $(\widehat{\boldsymbol{\alpha}}_{\text{sp}}, \widehat{G}_{\text{sp}})$ *is used to generate the bootstrap data $X_1^*, \ldots, X_n^*$, we have*

$$d_K\big(\mathcal{L}^*\big(\sqrt{n_m}\big(T_n^* - f\big(\boldsymbol{\xi}^*\big)\big)\big), \mathcal{L}\big(\sqrt{n_m}\big(T_n - f(\boldsymbol{\xi})\big)\big)\big) = o_P(1) \tag{3.9}$$

as $n \to \infty$, where $T_n^* = T_n(X_1^*, \ldots, X_n^*)$ and $\boldsymbol{\xi}^* = E^*(\boldsymbol{g}(X_t^*, \ldots, X_{t+m-1}^*))$.

Note that in comparison to Corollary 3.4, the bootstrap consistency result in Corollary 3.7 holds without the parametric restriction of Assumption 4.

Remark 3.8 (On moment restrictions and INAR order selection).

(i) At a first glance, the moment requirement in Corollary 3.7 may sound quite restrictive. But in a count data context, the non-existence of moments is usually not a topic. Most count models have existing moments up to any order, and also in applications, one will hardly find "heavy-tailed" count data.

(ii) Another possible issue for practice is the selection of the INAR's model order p. But since the autocorrelation structures of AR(p) and INAR(p) models coincide, order selection for INAR models can be done as in the usual AR(p) case, i.e., by looking at the sample partial autocorrelation function, or by using information criteria as for example, AIC, BIC, etc. As can be seen from our subsequent simulation study, an underfitting of the DGP should be avoided. Thus, one should better choose a somewhat larger model order in case of doubt, although this goes along with increased computational costs.

4. Simulation study and data example

4.1. Description of simulation study

We investigated the performance of the discussed bootstrap procedure in an extensive simulation study, where we analyzed the true coverage of 95% confidence intervals for diverse statistics. To compute a bootstrap confidence interval for an estimator $\widehat{\theta}$, we first computed an appropriate centering $\mathrm{cent}(\widehat{\theta}^*)$ (see Appendix E.2 for details), and then the centered bootstrap estimates $\widehat{\theta}^*_{\mathrm{cent}} := \widehat{\theta}^* - \mathrm{cent}(\widehat{\theta}^*)$. Using the $(1-\alpha/2)$- and $\alpha/2$-quantiles from these centered bootstrap estimates, the bootstrap confidence interval is defined as

$$\big[\widehat{\theta} - q_{1-\alpha/2}\big(\widehat{\theta}^*_{\mathrm{cent}}\big);\, \widehat{\theta} - q_{\alpha/2}\big(\widehat{\theta}^*_{\mathrm{cent}}\big)\big].$$

In our simulations, we considered the following estimators, which refer to important characteristics of a count process (Weiß, Puig and Homburg [53]):

- $\overline{X} := \frac{1}{n}\sum_{t=1}^n X_t$; estimator of the observations' mean μ_X;
- $\widehat{\gamma}(0) := \frac{1}{n}\sum_{t=1}^n (X_t - \overline{X})^2$; estimator of the observations' variance σ_X^2;
- $\widehat{I}_{\mathrm{disp}} := \widehat{\gamma}(0)/\overline{X}$; estimator of the observations' dispersion index σ_X^2/μ_X;
- $\widehat{\gamma}(1) := \frac{1}{n}\sum_{t=2}^n (X_t - \overline{X})(X_{t-1} - \overline{X})$; estimator of the first-order autocovariance $\gamma_X(1)$;
- $\widehat{\rho}(1) := \widehat{\gamma}(1)/\widehat{\gamma}(0)$; estimator of the first-order autocorrelation $\rho_X(1)$;
- $\overline{X}(1-\widehat{\rho}(1))$; estimator for $\mu_X(1-\rho_X(1))$, a quantity, which coincides with the innovations' mean μ_ϵ for autoregressive order $p=1$;
- $\widehat{p}_0 := \frac{1}{n}\sum_{t=1}^n \mathbb{1}(X_t = 0)$; estimator of the probability p_0 for observing a zero;
- $\widehat{I}_{\mathrm{z.i.}} := \ln(\widehat{p}_0)/\overline{X} + 1$; estimator of the zero inflation index;
- $\widehat{I}_{\mathrm{z.m.}} := \widehat{p}_0 \exp(\overline{X}) - 1$; modified type of zero index that can also handle the case of observing no zeros.

To be able to evaluate the performance of the proposed semi-parametric INAR-Bootstrap (spINAR-Bootstrap), several types of benchmarks were considered: the parametric Poisson INAR-Bootstrap (pINAR-Bootstrap, that is, which assumes an underlying INAR process having Poisson-distributed innovations), the Circular Block Bootstrap (CBB), the AR-Bootstrap (AR-Bootstrap), and the Markov Bootstrap (MB). For autoregressive order $p=1$, we also considered an asymptotic approximation as a further benchmark, that is, we utilized the asymptotic distributions for the estimators assuming an underlying Poisson INAR(1) process (see Appendix E.3 for details). Note that the AR bootstrap, which assumes a continuous AR process, is certainly not able to deal with a zero frequency. As an ad hoc approximation, we used the frequency of the event "$X_t < 0.5$" as the "zero frequency", also see Appendix E.2.

All bootstrap procedures were applied by once assuming autoregressive order $p=1$ (or Markov order $p=1$, respectively), and $p=2$ for the other time. When assuming $p=1$, we also considered the asymptotic approximation according to Appendix E.3, as mentioned before. For both cases, we worked with either Poisson innovations (as required by the parametric INAR bootstrap) or negative binomial (NB) innovations (this case constitutes a robustness check for the parametric INAR bootstrap). Furthermore, for $p=1$, we also used a Poisson INAR(2) model as the true data generating mechanism, thus constituting a robustness check for both INAR(1) bootstraps as well as for the AR(1) and Markov(1) bootstrap.

When generating Poisson innovations, we chose $\mu_\epsilon \in \{1, 2.5\}$. For NB$(n, \pi)$-innovations, we took $\mu_\epsilon \in \{1, 2.5\}$ and $\frac{\sigma_\epsilon^2}{\mu_\epsilon} \in \{1.5, 2.0\}$, and the parameters n, π were computed from the relations $\frac{\sigma_\epsilon^2}{\mu_\epsilon} = \frac{1}{\pi}$ and $\mu_\epsilon = n\frac{1-\pi}{\pi}$. When generating an INAR(1) process, the thinning parameter α (which coincides with $\rho(1)$ in this case) was taken as either 0.4 or 0.8. To keep a generated INAR(2) process comparable, we fixed the values $\rho(1) \in \{0.4, 0.8\}$ and $\alpha_2 \in \{0.15, 0.3\}$, and we computed α_1 from the relation $\rho(1) = \frac{\alpha_1}{1-\alpha_2}$. The considered sample sizes are $n \in \{100, 250, 500, 1000\}$. For each parametrization, 500 sample paths were simulated, and for each sample path, every bootstrap loop ran through 500 replications. For cases where the stationary distribution was not directly available, a pre-run of length 100 was used to approximately reach the steady state.

All computations were done using MATLAB. Technical details concerning the semi-parametric estimation procedure of Drost, van den Akker and Werker [13] (like starting values, numerical optimization routine, etc.) are summarized in Appendix E.1. When applying the circular block bootstrap, the block length was selected using the MATLAB package `opt_block_length_REV_dec07.m` offered by Andrew Patton,[2] which implements a corrected version of the automatic block-length selection procedure by Politis and White [45], Patton, Politis and White [43]. For the Markov bootstrap, we used the approach described in Section 3 of Basawa, Green, McCormick and Taylor [3] to avoid a degenerate estimated transition matrix as it could be caused by unobserved pairs or triples of states.

4.2. Discussion of simulation results

This section summarizes our main findings from the simulation study. Since it is not possible to print the complete tables, we shall illustrate our findings by only small excerpts, but the complete tables are provided as a Supplementary File at Jentsch and Weiß [23].

Let us first compare both types of INAR bootstraps. If the underlying process is Poisson INAR(1), then, as expected, the parametric INAR(1) bootstrap (which also assumes such a Poisson model) is usually superior to the semi-parametric one (see Tables 1 and 2 for illustration).

Table 1. INAR(1) process with $\mu_\epsilon = 1$, $\alpha = 0.8$; coverage for $\mu_X = 5$

	Poi-INAR(1), i.e., $\frac{\sigma_\epsilon}{\mu_\epsilon} = 1$						NB-INAR(1) with $\frac{\sigma_\epsilon}{\mu_\epsilon} = 1.5$					
n	spINAR	pINAR	CBB	AR	MB	asymp	spINAR	pINAR	CBB	AR	MB	asymp
100	0.904	0.890	0.816	0.870	0.816	0.926	0.910	0.868	0.798	0.868	0.802	0.874
250	0.930	0.936	0.862	0.914	0.880	0.944	0.934	0.888	0.868	0.902	0.850	0.888
500	0.938	0.928	0.898	0.924	0.892	0.934	0.946	0.918	0.914	0.934	0.902	0.926
1000	0.948	0.950	0.920	0.946	0.916	0.954	0.940	0.908	0.918	0.948	0.926	0.914

[2] http://public.econ.duke.edu/~ap172/

Table 2. INAR(1) process with $\mu_\epsilon = 1$, $\alpha = 0.8$; coverage for $\sigma_X^2 = 5$ resp. $\sigma_X^2 \approx 6.389$

	Poi-INAR(1), i.e., $\frac{\sigma_\epsilon}{\mu_\epsilon} = 1$						NB-INAR(1) with $\frac{\sigma_\epsilon}{\mu_\epsilon} = 1.5$					
n	spINAR	pINAR	CBB	AR	MB	asymp	spINAR	pINAR	CBB	AR	MB	asymp
100	0.804	0.858	0.708	0.762	0.716	0.906	0.812	0.726	0.650	0.726	0.626	0.798
250	0.874	0.914	0.794	0.842	0.760	0.938	0.880	0.798	0.780	0.838	0.758	0.812
500	0.898	0.918	0.848	0.870	0.818	0.932	0.922	0.818	0.840	0.884	0.800	0.834
1000	0.934	0.936	0.878	0.912	0.870	0.944	0.952	0.848	0.890	0.918	0.858	0.864

In fact, the asymptotic CIs do even better than the pINAR-bootstrap in this case. This is not surprising, since the considered asymptotic distributions have been proven to constitute very good approximations for the true distribution of the statistics (see the references in Appendix E.3). But if the true data generating mechanism is a non-Poisson INAR(1) process (like the NB-INAR(1) process in Tables 1 and 2), then both the pINAR-bootstrap as well as the asymptotic approach generally become worse than the spINAR-bootstrap. So these methods are not particularly robust to violations of the assumed distribution, restricting their applicability for real data. The coverages for pINAR and asymp degrade, since these methods are not able to reproduce the increased variance of the innovations. The bootstrap CIs according to CBB, AR and MB are robust w.r.t. the innovations' distribution, with CBB and MB converging only slowly to the nominal confidence level of 95%. The AR bootstrap works fine if computing a CI for the marginal mean μ_X (see Table 1), but it is not able to produce a reliable CI, for example, for the marginal variance σ_X^2 (see Table 2). The latter is not surprising in view of our earlier discussion in Section 2 (in particular, Lemma 2.1 and Example 2.2).

It is also plausible that the (continuous) AR bootstrap does not produce reliable confidence intervals for the zero probability (or the related zero indices): for $\widehat{I_{\text{z.i.}}}$, the coverages sometimes even decrease with increasing sample size, while in other cases, the coverages tend towards 1 instead of the nominal level 0.95 (see Table 3 for illustration). At this point, it has to be mentioned that CIs for zero statistics are sometimes not applicable (highlighted accordingly in the complete tables); for a Poisson INAR(1) process with $\mu_\epsilon = 2.5$, $\alpha = 0.8$, for instance, the true probability for observing a zero equals $\exp(-\mu_\epsilon/(1-\alpha)) \approx 3.73 \cdot 10^{-6}$, that is, one will hardly ever observe a zero at all.

Table 3. INAR(1) process with $\mu_\epsilon = 1$, $\alpha = 0.8$; coverage for $p_0 \approx 0.007$ resp. $p_0 \approx 0.012$

	Poi-INAR(1), i.e., $\frac{\sigma_\epsilon}{\mu_\epsilon} = 1$						NB-INAR(1) with $\frac{\sigma_\epsilon}{\mu_\epsilon} = 1.5$					
n	spINAR	pINAR	CBB	AR	MB	asymp	spINAR	pINAR	CBB	AR	MB	asymp
100	0.524	0.604	0.324	0.750	0.324	0.970	0.596	0.484	0.432	0.718	0.432	0.872
250	0.708	0.762	0.666	0.912	0.670	0.956	0.750	0.578	0.584	0.896	0.586	0.828
500	0.806	0.842	0.676	0.990	0.676	0.962	0.880	0.692	0.778	0.982	0.772	0.820
1000	0.922	0.944	0.780	1.000	0.782	0.962	0.930	0.764	0.818	0.994	0.796	0.828

Table 4. INAR(2) process with $\mu_\epsilon = 1$, $\alpha_1 = 0.68$, $\alpha_2 = 0.15$; coverage for $\mu_X \approx 5.882$

	Falsely assuming $p = 1$						Correctly assuming $p = 2$		
n	spINAR	pINAR	CBB	AR	MB	asymp	spINAR	pINAR	AR
100	0.758	0.776	0.770	0.818	0.746	0.796	0.834	0.860	0.850
250	0.786	0.798	0.842	0.848	0.810	0.808	0.884	0.896	0.884
500	0.810	0.822	0.888	0.886	0.836	0.822	0.936	0.940	0.940
1000	0.830	0.850	0.926	0.914	0.892	0.860	0.956	0.954	0.950

The MB bootstrap makes use of the INAR's Markov property and is therefore a reasonable candidate at the first glance. However, it severely suffers from the problem that some states (or combinations thereof) are not observed in a given time series of finite length. Our simulations show that this problem becomes increasingly severe with decreasing sample size n on the one hand, and with increasing mean μ_ϵ and dispersion ratio $\frac{\sigma_\epsilon}{\mu_\epsilon}$ on the other hand (then it becomes more and more probable to not observe all values between 0 and $m := \max\{X_1, \ldots, X_n\}$). In particular, while the Markov(1) bootstrap (with order $p = 1$, that is, with $(m+1)m$ parameters) works reasonably well at least in the Poisson case ($\frac{\sigma_\epsilon}{\mu_\epsilon} = 1$) with the small mean $\mu_\epsilon = 1$, the Markov(2) bootstrap (with $(m+1)^2 m$ parameters) showed extremely bad coverages; for this reason, we do not consider the MB bootstrap anymore in the sequel.

Among the remaining schemes, the CBB bootstrap often produces the worst coverages, which is reasonable since it is the only fully non-parametric procedure. The semi-parametric spINAR-bootstrap, in contrast, which only assumes an INAR(p) structure but is non-parametric in the innovations' distribution, usually produces much better coverage rates. There is only one important exception: If the true data generating process is (Poisson) INAR(2), but the spINAR-, pINAR- and AR-bootstraps (also the asymptotic CIs and the MB bootstrap) assume an autoregressive order $p = 1$, then the CBB bootstrap (where we do not need to specify p at all, but a block length) shows the best coverages, while the before-mentioned approaches are not robust against such a misspecified model order. This is illustrated by the results shown in Tables 4 (mean μ_X) and 5 (variance σ_X^2), where the autoregressive parameters α_1, α_2 are chosen such that still $\rho(1) = 0.8$, as in the previous tables.

Table 5. INAR(2) process with $\mu_\epsilon = 1$, $\alpha_1 = 0.68$, $\alpha_2 = 0.15$; coverage for $\sigma_X^2 \approx 8.610$

	Falsely assuming $p = 1$						Correctly assuming $p = 2$		
n	spINAR	pINAR	CBB	AR	MB	asymp	spINAR	pINAR	AR
100	0.612	0.598	0.634	0.666	0.564	0.658	0.720	0.754	0.682
250	0.680	0.670	0.774	0.782	0.712	0.696	0.822	0.844	0.798
500	0.658	0.640	0.806	0.774	0.706	0.662	0.872	0.894	0.816
1000	0.696	0.678	0.850	0.818	0.792	0.690	0.922	0.920	0.854

Table 6. NB-INAR(2) process with $\mu_\epsilon = 1$, $\alpha_1 = 0.68$, $\alpha_2 = 0.15$, $\frac{\sigma_\epsilon}{\mu_\epsilon} = 1.5$

	Coverage for $\mu_X \approx 5.882$				Coverage for $\sigma_X^2 \approx 10.031$			
n	spINAR	pINAR	CBB	AR	spINAR	pINAR	CBB	AR
100	0.810	0.816	0.754	0.828	0.702	0.658	0.628	0.664
250	0.896	0.872	0.834	0.892	0.842	0.792	0.784	0.798
500	0.920	0.896	0.872	0.924	0.904	0.844	0.828	0.844
1000	0.944	0.920	0.896	0.922	0.910	0.864	0.872	0.874

To continue with the INAR(2) case, let us also have a look at Table 6, where again coverages for mean and variance are shown, but now for an NB-INAR(2) process (with correctly specified $p = 2$). Comparing with the right parts in Tables 4 and 5, we see again that the coverages for pINAR degrade, in analogy to the above discussion (Tables 1 and 2) for autoregressive order $p = 1$, while spINAR is robust with respect to the marginal distribution.

4.3. Real data application

Jung and Tremayne [27] analyzed a time series (length $n = 1632$) of counts of so-called iceberg orders concerning the Lufthansa stock traded in the XETRA system of Deutsche Börse. The time series gives the number of iceberg orders (for the buy side) per 10 minutes for 32 trading days, and the obtained counts range between 0 and 7. An analysis of the (partial) autocorrelation function (Jung and Tremayne [27], Figure 3) shows that a second-order autoregressive model seems to be appropriate for describing these data (with $\widehat{\rho}(1) \approx 0.507$ and $\widehat{\rho}(2) \approx 0.397$). Sample mean and variance of these data are given by about 0.778 and 0.819, respectively, that is, the observations are nearly equidispersed. The zero frequency equals about 0.463, which is rather close to the corresponding Poisson value $\exp(-\overline{x}) \approx 0.460$. Further descriptive statistics are provided by Appendix E.4. There, also the obtained estimates are shown, when semi-parametrically fitting an INAR(1) or INAR(2) model. Any bootstrap procedure was done with 10 000 replications. The obtained 95% confidence intervals are summarized in Appendix E.4.

Since we do not know the true model behind the iceberg counts, interpretations have to be done with caution. Generally, the length of the CIs increases if the bootstrap procedures “spINAR”, “pINAR” and “AR” were run based on a 2nd-order model instead of a 1st-order one, which is plausible in view of the need for estimating the additional parameter α_2. For a similar reason, it is plausible that the “spINAR” intervals are usually larger than the “pINAR” intervals. For $p = 1$, the “pINAR” intervals are nearly identical to the asymptotic ones, which is reasonable since both approaches assume an underlying Poisson INAR(1) model. The intervals obtained by “spINAR” with $p = 2$ are often close to those obtained by “CBB”. The “AR” intervals are sometimes similar to the other intervals, for example, for the mean $\overline{x}$, but show strong deviations in other cases. For example, the AR-CIs with $p = 2$ for $\widehat{\gamma}(0)$, $\widehat{\gamma}(1)$ are much more narrow than the corresponding spINAR- and pINAR-CIs. Some common conclusions, which are implied by any of the respective CI approaches, are non-significant deviations of $\widehat{I}_{\text{disp}}$ from 1, and of $\widehat{I}_{\text{z.i.}}$, $\widehat{I}_{\text{z.m.}}$ from 0, that is, neither a significant degree of over-/underdispersion nor of zero inflation/deflation is established for the observations.

5. Conclusion

Although the INAR(p) models share the autocorrelation structure with the AR(p) models, the traditional AR bootstrap is generally not consistent if applied to an INAR process. Therefore, we proposed a general INAR bootstrap scheme, which we proved to be consistent (under mild conditions) for a class of functions of generalized means. In particular, this bootstrap consistency applies to the considered parametric INAR bootstrap, if such parametric assumptions hold for the DGP, as well as to the novel semi-parametric INAR bootstrap without further parametric assumptions. The simulation study concerning bootstrap confidence intervals showed that this semi-parametric INAR bootstrap is very promising for applications, as it showed a good performance for the different model parametrizations and innovations' distributions considered for our analyses. Therefore, future research should investigate further applications of the novel semi-parametric INAR bootstrap, for example, for the bias correction of estimators, or for diagnostic tests of the INAR model structure.

Appendix A: Proofs

A.1. Proof of Theorem 3.2

The basic structure of the proof resembles the proof of Theorem 3.3 in Bühlmann [8]. For more details, we refer to his technical report (Bühlmann [7]) and to the proof of Theorem 4.2 in Meyer, Jentsch and Kreiß [40], where a corresponding result is proved for random fields. The main arguments are the same here, but we have to address the non-linear and discrete-valued structure of INAR processes caused by the randomness of the binomial thinning operations in the model equations (1.3). In particular, this randomness inherent to the recursive autoregressive structure of INAR processes requires different arguments. We will sketch the main steps of the proof in the following.

First, let $\boldsymbol{Y}_t = (X_t, \ldots, X_{t+m-1})'$ and $\boldsymbol{Y}_t^* = (X_t^*, \ldots, X_{t+m-1}^*)'$ denote sub-sequences of the original INAR process $(X_t, t \in \mathbb{Z})$ and the corresponding bootstrap process $(X_t^*, t \in \mathbb{Z})$. The general strategy to show bootstrap consistency for the statistic T_n is to neglect the function f and to prove that

$$\frac{1}{\sqrt{n_m}} \sum_{t=1}^{n_m} \big(\boldsymbol{g}(\boldsymbol{Y}_t) - E\big(\boldsymbol{g}(\boldsymbol{Y}_t)\big)\big) \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{A.1}$$

as well as

$$\frac{1}{\sqrt{n_m}} \sum_{t=1}^{n_m} \big(\boldsymbol{g}\big(\boldsymbol{Y}_t^*\big) - E^*\big(\boldsymbol{g}\big(\boldsymbol{Y}_t^*\big)\big)\big) \xrightarrow{d} \mathcal{N}(0, \Sigma) \tag{A.2}$$

in probability, where

$$\Sigma := \sum_{h=-\infty}^{\infty} \operatorname{Cov}\big(\boldsymbol{g}(\boldsymbol{Y}_h), \boldsymbol{g}(\boldsymbol{Y}_0)\big).$$

Under the smoothness conditions in Assumption 1, it is straightforward to treat the function f by using the Delta method such that it suffices to show (A.1) and (A.2). As the proof of (A.1) follows by the same (and simpler) arguments, we shall execute only the proof of (A.2) in the following.

From Lemma B.2, we know that the INAR bootstrap process has a corresponding INMA(∞) representation

$$X_t^* = \sum_{k=0}^{\infty} \boldsymbol{u}'\big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}^*.$$

We truncate the INMA(∞) representation of X_t^* and consider $X_t^{*(q)}$, $q \in \mathbb{N}_0$, in the following, where

$$X_t^{*(q)} = \sum_{k=0}^{q} \boldsymbol{u}'\big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}^*,$$

see also the proof of Lemma B.2. The random vectors $\boldsymbol{Y}_t^{(q)}$ and $\boldsymbol{Y}_t^{(q)*}$ are defined in an analogous way as $\boldsymbol{Y}_t$ and $\boldsymbol{Y}_t^*$ above.

Since $\boldsymbol{g}$ is differentiable with hth derivative fulfilling a Lipschitz condition by Assumption 1, a Taylor series expansion of order h and Minkowski's inequality allows to show that

$$E^*\big(\big|g_u\big(\boldsymbol{Y}_t^{(q)*}\big)\big|^{2+2/(h+1)}\big) = O_P(1) \qquad \text{for all } 1 \le u \le d \tag{A.3}$$

holds, where we have to use that $E^*(\epsilon_t^{*2(h+2)}) = O_P(1)$ from Assumption 3(ii) implies that $E^*(X_t^{*2(h+2)}) = O_P(1)$ by Lemma B.2. For details, we refer to Meyer, Jentsch and Kreiß [39], proofs of equation (7.5) and Lemma 7.2, and Bühlmann [7], page 22.

To continue with the proof, note that the random variable $X_t^{*(q)}$ is composed of $\{\epsilon_t^*, \dots, \epsilon_{t-q}^*\}$ and $\{\widehat{\mathbf{A}}\circ_t^*, \dots, \widehat{\mathbf{A}}\circ_{t-q+1}^*\}$. As $\{\epsilon_t^*, t \in \mathbb{Z}\}$ and $\{\widehat{\mathbf{A}}\circ_t^*, t \in \mathbb{Z}\}$ are i.i.d. random variables and operations, respectively, we have that $\{X_t^{*(q)}\}$ is a q-dependent process. Using a truncation argument for the function $\boldsymbol{g}$ as in Bühlmann [7], page 21, this allows us to show

$$\operatorname{Cov}^*\left(\frac{1}{\sqrt{n_m}}\sum_{t=1}^{n_m} g_u\big(\boldsymbol{Y}_t^{*(q)}\big), \frac{1}{\sqrt{n_m}}\sum_{t=1}^{n_m} g_v\big(\boldsymbol{Y}_t^{*(q)}\big)\right) = \big(\widetilde{\Sigma}^{(q)}\big)_{uv} + o_P(1) \qquad \text{for all } 1 \le u, v \le d,$$

where, for $\underline{\widetilde{\mathbf{Y}}}_t^{(q)} = (\widetilde{X}_t^{(q)}, \dots, \widetilde{X}_{t+m-1}^{(q)})'$,

$$\widetilde{\Sigma}^{(q)} := \sum_{h=-(q+m-1)}^{q+m-1} \operatorname{Cov}\big(\boldsymbol{g}\big(\underline{\widetilde{\mathbf{Y}}}_h^{(q)}\big), \boldsymbol{g}\big(\underline{\widetilde{\mathbf{Y}}}_0^{(q)}\big)\big).$$

Here, $(\widetilde{X}_t^{(q)}, t \in \mathbb{Z})$ is the truncated INMA(q) process corresponding to the INMA(∞) representation of the companion INAR process $(\widetilde{X}_t, t \in \mathbb{Z})$ in (3.5). This INMA(∞) representation is assured to exist again by Lemma B.2. Now, the Cramér–Wold device is applied to show the convergence of

$$\frac{1}{\sqrt{n_m}} \sum_{t=1}^{n_m} \big(\boldsymbol{c}'\boldsymbol{g}\big(\boldsymbol{Y}_t^{*(q)}\big) - E^*\big(\boldsymbol{c}'\boldsymbol{g}\big(\boldsymbol{Y}_t^{*(q)}\big)\big)\big)$$

for $\boldsymbol{c} = (c_1, \dots, c_d)' \in \mathbb{R}^d$, and we define $\ell(\boldsymbol{x}) = \boldsymbol{c}'\boldsymbol{g}(\boldsymbol{x}) = \sum_{u=1}^d c_u g_u(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^m$. The q-dependence of the bootstrap INAR process $(X_t^{*(q)})$ allows to use a big-block–small-block argument together with Lindeberg's CLT for triangular arrays. Here, it remains to show the Lindeberg condition, which can be proved using the same arguments as in Bühlmann [7], top of page 22. Here, to make use of the bound from Yokoyama [55] for sums of strictly stationary processes, note that the process $(X_t^{*(q)})$ satisfies this property for each fixed $n \in \mathbb{N}$ and that, analogous to (A.3),

$$E^*\big(\big|\ell\big(\boldsymbol{Y}_t^{(q)*}\big)\big|^{2+2/(h+1)}\big) = O_P(1)$$

holds, which altogether proves asymptotic normality and we get

$$\frac{1}{\sqrt{n_m}} \sum_{t=1}^{n_m} \big(\boldsymbol{g}\big(\boldsymbol{Y}_t^{*(q)}\big) - E^*\big(\boldsymbol{g}\big(\boldsymbol{Y}_t^{*(q)}\big)\big)\big) \xrightarrow{d} \mathcal{N}\big(0, \widetilde{\Sigma}^{(q)}\big) \tag{A.4}$$

in probability. Using again the smoothness properties of $\boldsymbol{g}$, a Taylor series expansion and Minkowski's inequality, one can show, as $q \to \infty$, that

$$\left| \sum_{h=-(q+m-1)}^{q+m-1} \operatorname{Cov}\big(\boldsymbol{g}\big(\underline{\widetilde{\mathbf{Y}}}_h^{(q)}\big), \boldsymbol{g}\big(\underline{\widetilde{\mathbf{Y}}}_0^{(q)}\big)\big) - \sum_{h=-(q+m-1)}^{q+m-1} \operatorname{Cov}\big(\boldsymbol{g}(\underline{\widetilde{\mathbf{Y}}}_h), \boldsymbol{g}(\underline{\widetilde{\mathbf{Y}}}_0)\big) \right| = o(1)$$

and

$$\left| \sum_{|h|>q+m-1} \operatorname{Cov}\big(\boldsymbol{g}(\underline{\widetilde{\mathbf{Y}}}_h), \boldsymbol{g}(\underline{\widetilde{\mathbf{Y}}}_0)\big) \right| = o(1).$$

Following the truncation argument in Bühlmann [7], page 21, and using Lemma B.3, this leads to $\lim_{q\to\infty} \widetilde{\Sigma}^{(q)} = \Sigma$. Now, to conclude the proof of (A.2), it remains to show that truncating the INMA(∞) process X_t^* to get $X_t^{*(q)}$ is negligible. By exploiting the q-dependence of $(X_t^{*(q)})$, this can be done similarly to Bühlmann [7], pages 23, 24, using again a suitable Taylor series expansion of $\ell(\cdot)$; see also Meyer, Jentsch and Kreiß [40], proof of Theorem 4.2.

A.2. Proof of Corollary 3.4

To make use of Theorem 3.2, the goal in the following is to show that the parametric Assumption 4 together with the parametric INAR bootstrap implementation implies Assumption 3.

First of all, Yule–Walker estimates $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ in Step 1 of the parametric INAR bootstrap scheme are known to satisfy Assumption 3(i). To show part (ii), first, we get easily

$$d_1(\widehat{G}, G) = d_1(G_{\widehat{\boldsymbol{\theta}}}, G_{\boldsymbol{\theta}_0}) = \sum_{k=0}^{\infty} \big| G_{\widehat{\boldsymbol{\theta}}}(k) - G_{\boldsymbol{\theta}_0}(k) \big| = o_P(1) \tag{A.5}$$

from the $\sqrt{n}$-consistency $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1)$ and the continuity of $\boldsymbol{\theta} \mapsto G_{\boldsymbol{\theta}}$. Similarly, we prove next that Assumption 3(ii) holds, i.e. $E^*_{G_{\widehat{\boldsymbol{\theta}}}}(\epsilon_t^{*s}) \to E_G(\epsilon_t^s) < \infty$ in probability for $s = 1, \ldots, 2(h+2)$ with h specified by Assumption 1. Now, let $\epsilon, \eta > 0$. Then, for all $M \in \mathbb{N}$, we have

$$\begin{aligned}
&P\big(\big|E^*_{G_{\widehat{\boldsymbol{\theta}}}}\big(\varepsilon_t^{*s}\big) - E_G\big(\epsilon_t^s\big)\big| \geq \varepsilon\big) \\
&\quad \leq P\left(\sum_{k=0}^{M} k^s \big|G_{\widehat{\boldsymbol{\theta}}}(k) - G_{\boldsymbol{\theta}_0}(k)\big| \geq \varepsilon/3\right) + P\left(\sum_{k=M+1}^{\infty} k^s \big|G_{\widehat{\boldsymbol{\theta}}}(k)\big| \geq \varepsilon/3\right) \\
&\qquad + P\left(\sum_{k=M+1}^{\infty} k^s \big|G_{\boldsymbol{\theta}_0}(k)\big| \geq \varepsilon/3\right) \\
&\quad \leq P\left(\sum_{k=0}^{\infty} \big|G_{\widehat{\boldsymbol{\theta}}}(k) - G_{\boldsymbol{\theta}_0}(k)\big| \geq \varepsilon/\big(3M^s\big)\right) \\
&\qquad + P\left(\sum_{k=M+1}^{\infty} k^s \big|G_{\widehat{\boldsymbol{\theta}}}(k)\big| \geq \varepsilon/3, |\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < \delta\right) \\
&\qquad + P\big(|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| \geq \delta\big) + P\left(\sum_{k=M+1}^{\infty} k^s \big|G_{\boldsymbol{\theta}_0}(k)\big| \geq \varepsilon/3\right) \\
&\quad = I + II + III + IV.
\end{aligned}$$

By Assumption 4, term II can be bounded by

$$P\left(\sup_{\boldsymbol{\theta} \in B_{\boldsymbol{\theta}_0,\delta}} \sum_{k=M+1}^{\infty} k^s \big|G_{\boldsymbol{\theta}}(k)\big| \geq \varepsilon/3, |\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < \delta\right) \leq P\left(\sup_{\boldsymbol{\theta} \in B_{\boldsymbol{\theta}_0,\delta}} \sum_{k=M+1}^{\infty} k^s \big|G_{\boldsymbol{\theta}}(k)\big| \geq \varepsilon/3\right)$$

Now, we can choose M large enough such that the latter term as well as term IV equal zero. Further, from the smoothness condition in Assumption 4 and the $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\theta}}$, we can choose $n_0 = n_0(M, \varepsilon, \eta, \delta)$ large enough such that $I \leq \eta/2$ and $III \leq \eta/2$. Altogether this leads to

$$P\big(\big|E^*_{G_{\widetilde{\boldsymbol{\theta}}}}\big(\varepsilon_t^{*s}\big) - E_G\big(\epsilon_t^s\big)\big| \geq \varepsilon\big) \leq \eta.$$

A.3. Proof of Corollary 3.7

We make use of Theorem 3.2 and show that the semi-parametric INAR bootstrap implementation based on the estimation approach of Drost, van den Akker and Werker [13] implies Assumption 3.

First of all, note that Theorem 2 in Drost, van den Akker and Werker [13] holds under the assumptions of this corollary. Then, we get immediately from the first part of (3.8) that the coefficient estimates $\widehat{\alpha}_1, \dots, \widehat{\alpha}_p$ in Step 1 of the semi-parametric INAR bootstrap scheme satisfy Assumption 3(i). Further, from the second part of (3.8), we get also immediately the main prerequisites to show also part (ii) of Assumption 3, i.e. $E^*_{\widehat{G}}(\epsilon_t^{*s}) \to E_G(\epsilon_t^s) < \infty$ in probability for $s = 1, \dots, 2(h+2)$ in probability with h specified by Assumption 1. From $\widehat{G}(k) = 0$ for $k > M_n := \max\{X_1, \dots, X_n\}$, we get

$$\begin{aligned}
\left|E^*_{\widehat{G}_{\mathrm{sp}}}\left(\epsilon_t^{*s}\right) - E_G\left(\epsilon_t^s\right)\right| &= \left|\sum_{k=0}^{\infty} k^s \widehat{G}(k) - \sum_{k=0}^{\infty} k^s G(k)\right| = \left|\sum_{k=0}^{M_n} k^s \widehat{G}(k) - \sum_{k=0}^{\infty} k^s G(k)\right| \\
&\leq \sum_{k=0}^{M_n} k^s \left|\widehat{G}(k) - G(k)\right| + \sum_{k=M_n+1}^{\infty} k^s G(k) \\
&\leq M_n^s \left(\sum_{k=0}^{M_n} \left|\widehat{G}(k) - G(k)\right|\right) + \sum_{k=M_n+1}^{\infty} k^s G(k) \\
&=: A_1 + A_2.
\end{aligned}$$

If the support of $\{\epsilon_t\}$ is bounded, then $A_2 = 0$ after one observation X_t attains a value greater or equal to the largest possible innovation, which happens with probability tending to one. If the support is unbounded, $M_n \to \infty$ such that $A_2 = o(1)$ as we have $E_G(\epsilon_t^s) < \infty$ by Assumption 2, i.e., $(k^s G(k))_k$ is summable. Concerning the first summand, we make use of the second part of (3.8). This leads to

$$A_1 = \frac{M_n^s}{\sqrt{n}} \left(\sqrt{n} \sum_{k=0}^{M_n} \left|\widehat{G}(k) - G(k)\right|\right) \leq \frac{M_n^s}{\sqrt{n}} \left(\sqrt{n} \sum_{k=0}^{\infty} \left|\widehat{G}(k) - G(k)\right|\right) = O_P\left(\frac{M_n^s}{\sqrt{n}}\right).$$

Further, for all $x, k \geq 0$, we get from the Markov inequality

$$\begin{aligned}
P\left(M_n^s/\sqrt{n} > x\right) &= P\left(M_n > x^{1/s} n^{1/(2s)}\right) = P\left(\bigcup_{i=1}^{n} \left\{X_i > x^{1/s} n^{1/(2s)}\right\}\right) \\
&\leq \sum_{i=1}^{n} P\left(X_i > x^{1/s} n^{1/(2s)}\right) = n P\left(X_1 > x^{1/s} n^{1/(2s)}\right) \\
&\leq n \frac{E(X_1^k)}{x^{k/s} n^{k/(2s)}} = \frac{E(X_1^k)}{x^{k/s}} n^{1-k/(2s)},
\end{aligned}$$

and the last right-hand side converges to zero if $k > 2s$ as well as $E(X_1^k) < \infty$. As we can choose $k > 4(h+2)$, this implies that $M_n^s/\sqrt{n} = o_P(1)$ holds for $s \leq 2(h+2)$, which concludes that Assumption 3(ii) is fulfilled.

Appendix B: Auxiliary results

Lemma B.1 (Basic properties of the INAR(p) **process).** *Under Assumption 2 with $s = 1$, the* INAR(p) *recursion (1.3) has a strictly stationary solution $(X_t, t \in \mathbb{Z})$ with $E(X_t) = \mu_\epsilon/(1 - \sum_{i=1}^p \alpha_i) < \infty$. Precisely, by introducing the notation $\boldsymbol{X}_t := (X_t, \dots, X_{t-p+1})'$, $\boldsymbol{\epsilon}_t = (\epsilon_t, 0, \dots, 0)'$, $\boldsymbol{u} := (1, 0, \dots, 0)'$ and*

$$\mathbf{A} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \cdots & \alpha_p \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

and entry-wise binomial thinning $\circ_t$, the INAR(p) *recursion (1.3) can be represented as a p-dimensional* INAR(1) *recursion*

$$\boldsymbol{X}_t = \mathbf{A} \circ_t \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t, \qquad t \in \mathbb{Z} \tag{B.1}$$

such that $X_t = \boldsymbol{u}'\boldsymbol{X}_t$ holds. The INAR(p) *process can be represented as an* INMA(∞) *process, that is,*

$$X_t = \sum_{k=0}^{\infty} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^t\big)\boldsymbol{\epsilon}_{t-k}, \qquad t \in \mathbb{Z}, \tag{B.2}$$

where we set $[\mathbf{A}\circ]_{j=s}^t := \mathbf{A} \circ_t \mathbf{A} \circ_{t-1} \cdots \mathbf{A}\circ_s$ for $s \leq t$ and $[\mathbf{A}\circ]_{j=t+1}^t := 1$ for the empty index set. The convergence of (B.2) is in L_1-sense and almost surely. If Assumption 2 holds for some $h \in \mathbb{N}$, i.e., $E(\epsilon_t^h) < \infty$, it holds $E(X_t^h) < \infty$ and the convergence of (B.2) holds in L_h-sense.

Proof. By using the model equation (B.1) recursively q-times, we get

$$\boldsymbol{X}_t = \sum_{k=0}^{q} \big([\mathbf{A}\circ]_{j=t-k+1}^t\big)\boldsymbol{\epsilon}_{t-k} + [\mathbf{A}\circ]_{j=t-q}^t \boldsymbol{X}_{t-q+1} = \boldsymbol{X}_t^{(q)} + [\mathbf{A}\circ]_{j=t-q}^t \boldsymbol{X}_{t-q+1}, \tag{B.3}$$

where $\boldsymbol{X}_t^{(q)} = \sum_{k=0}^q ([\mathbf{A}\circ]_{j=t-k+1}^t)\boldsymbol{\epsilon}_{t-k}$ and $X_t^{(q)} = \boldsymbol{u}'\boldsymbol{X}_t^{(q)}$. As $[\mathbf{A}\circ]_{j=t-q}^t \boldsymbol{X}_{t-q+1}$ is always a vector with non-negative entries, it suffices to show that

$$\begin{aligned} E\big(\big|X_t - X_t^{(q)}\big|\big) &\leq E\big(\big|\boldsymbol{X}_t - \boldsymbol{X}_t^{(q)}\big|_1\big) = E\big(\big|[\mathbf{A}\circ]_{j=t-q}^t \boldsymbol{X}_{t-q+1}\big|_1\big) \\ &= \mathbf{1}_p' E\big([\mathbf{A}\circ]_{j=t-q}^t \boldsymbol{X}_{t-q+1}\big) \to 0 \end{aligned} \tag{B.4}$$

as $q \to \infty$ to prove the L_1-convergence of (B.2), where $\mathbf{1}_p = (1, \dots, 1)'$ denotes the p-dimensional vector of ones and $|\cdot|_1$ denotes the 1-norm. By conditioning, we get with standard arguments

$$\begin{aligned}
&E\big([\mathbf{A}\circ]_{j=t-q}^{t} \boldsymbol{X}_{t-q+1}\big) \\
&\quad = E\big(E\big(\mathbf{A} \circ_t [\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1} \mid [\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1}\big)\big) \\
&\quad = \sum_{\boldsymbol{k} \in \mathbb{N}_0^p} E\big(\mathbf{A} \circ_t [\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1} \mid [\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1} = \boldsymbol{k}\big) \\
&\qquad \times P\big([\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1} = \boldsymbol{k}\big) \\
&\quad = \sum_{\boldsymbol{k} \in \mathbb{N}_0^p} E\big(\mathbf{A} \circ_t \boldsymbol{k} \mid [\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1} = \boldsymbol{k}\big) P\big([\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1} = \boldsymbol{k}\big) \qquad \text{(B.5)} \\
&\quad = \sum_{\boldsymbol{k} \in \mathbb{N}_0^p} E(\mathbf{A} \circ_t \boldsymbol{k}) P\big([\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1} = \boldsymbol{k}\big) \\
&\quad = \sum_{\boldsymbol{k} \in \mathbb{N}_0^p} \mathbf{A}\boldsymbol{k} P\big([\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1} = \boldsymbol{k}\big) \\
&\quad = \mathbf{A} E\big([\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1}\big).
\end{aligned}$$

By successive application of the above, we get $E([\mathbf{A}\circ]_{j=t-q}^{t} \boldsymbol{X}_{t-q+1}) = \mathbf{A}^{q+1} \mu_X \mathbf{1}_p$. As $\alpha_i \in (0, 1)$ such that $\sum_{i=1}^{p} \alpha_i < 1$, all eigenvalues of $\mathbf{A}$ have modulus less than one; see, e.g., Latour [36], Lemma 2.2. Hence, we get $\mu_X \mathbf{1}_p' \mathbf{A}^{q+1} \mathbf{1}_p \to 0$ as $q \to \infty$ such that (B.4) holds. By similar arguments as used above, we get also

$$E(X_t) = \boldsymbol{u}' \sum_{k=0}^{\infty} \mathbf{A}^k E(\boldsymbol{\epsilon}_{t-k}) = \boldsymbol{u}'(\mathbf{I}_{p \times p} - \mathbf{A})^{-1} \boldsymbol{u} \mu_\epsilon = \frac{\mu_\epsilon}{1 - \sum_{i=1}^{p} \alpha_i}.$$

To show almost sure convergence, let $N \subset \Omega$ be the set of $\omega \in \Omega$, where convergence of

$$\sum_{k=0}^{\infty} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^{t}(\omega)\big) \boldsymbol{\epsilon}_{t-k}(\omega)$$

does not hold true and suppose $P(N) > 0$. As

$$\left(\sum_{k=0}^{n} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^{t}(\omega)\big) \boldsymbol{\epsilon}_{t-k}(\omega), n \in \mathbb{N} \right)$$

is a monotonely increasing sequence, we have

$$\sum_{k=0}^{\infty} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^{t}(\omega)\big)\boldsymbol{\epsilon}_{t-k}(\omega) = \infty$$

for all $\omega \in N$. Further, on the one hand, we get by the monotone convergence theorem

$$\begin{aligned}\lim_{n\to\infty} E\left(\sum_{k=0}^{n} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}\right) &= E\left(\lim_{n\to\infty}\sum_{k=0}^{n} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}\right)\\ &= E\left(\sum_{k=0}^{\infty} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}\right).\end{aligned}$$

On the other hand, we have

$$E\left(\sum_{k=0}^{n} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}\right) = \sum_{k=0}^{n} \boldsymbol{u}'\mathbf{A}^k E(\boldsymbol{\epsilon}_{t-k}) \le \boldsymbol{u}'\sum_{k=0}^{\infty}\mathbf{A}^k E(\boldsymbol{\epsilon}_{t-k}) < \infty,$$

but this is a contradiction, because

$$E\left(\sum_{k=0}^{\infty} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}\right) \ge E\left(\left(\sum_{k=0}^{\infty} \boldsymbol{u}'\big([\mathbf{A}\circ]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}\right)\mathbb{1}(N)\right) = \infty P(N) = \infty.$$

This proves $P(N) = 0$ and consequently almost sure convergence of (B.2). In particular, (B.2) is strictly stationary and solves the INAR(p) recursion (1.3). By plugging-in, this can be seen from

$$\begin{aligned} X_t - \sum_{i=1}^{p} \alpha_i \circ_t X_{t-i} &= \boldsymbol{u}'(\boldsymbol{X}_t - \mathbf{A}\circ_t \boldsymbol{X}_{t-1})\\ &= \boldsymbol{u}'\left(\sum_{k=0}^{\infty}\big([\mathbf{A}\circ]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k} - \sum_{k=0}^{\infty}\mathbf{A}\circ_t\big([\mathbf{A}\circ]_{j=t-k}^{t-1}\big)\boldsymbol{\epsilon}_{t-1-k}\right)\\ &= \boldsymbol{u}'\left(\sum_{k=0}^{\infty}\big([\mathbf{A}\circ]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k} - \sum_{k=1}^{\infty}\big([\mathbf{A}\circ]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}\right) \qquad \text{(B.6)}\\ &= \boldsymbol{u}'\boldsymbol{\epsilon}_t\\ &= \epsilon_t.\end{aligned}$$

Now, let $E(\epsilon_t^h) < \infty$ for some $h \in \mathbb{N}$. To show finiteness of $E(X_t^h)$, we make use of the norm notation $\|X_t\|_h = (E(X_t^h))^{1/h}$. By taking the norm on both sides of (1.3) and using Minkowski's inequality, we get

$$\|X_t\|_h \leq \sum_{i=1}^{p} \|\alpha_i \circ_{t,i} X_{t-i}\|_h + \|\epsilon_t\|_h. \tag{B.7}$$

Note that (B.7) does always hold as all involved summands are non-negative. Now, we shall compute $\|\alpha_i \circ_{t,i} X_{t-i}\|_h$. We get by standard arguments

$$\|\alpha_i \circ_{t,i} X_{t-i}\|_h^h = E\big((\alpha_i \circ_{t,i} X_{t-i})^h\big) = \sum_{k=0}^{\infty} E\big((\alpha_i \circ_{t,i} k)^h\big) P(X_{t-i} = k)$$

and as $\alpha_i \circ_{t,i} k \sim \text{Bin}(k, \alpha_i)$, we can make use of equation (3.6) (note the typo there, where $p^r/(n-r)!$ has to be replaced by $p^j/(n-j)!$) and (1.55) in Johnson, Kemp and Kotz [24] to rewrite the last expectation above. We get

$$E\big((\alpha_i \circ_{t,i} k)^h\big) = \sum_{j=0}^{h} S(h, j) \frac{k!}{(k-j)!} \alpha_i^j = \sum_{j=0}^{h} S(h, j) \sum_{l=0}^{j} s(j, l) k^l \alpha_i^j, \tag{B.8}$$

where $S(h, j)$ and $s(j, l)$ denote the Stirling numbers of second and first kind, respectively. The right-hand side of (B.8) can be decomposed to get

$$\sum_{j=0}^{h-1} S(h, j) \sum_{l=0}^{j} s(j, l) k^l \alpha_i^j + \sum_{l=0}^{h-1} s(h, l) k^l \alpha_i^h + k^h \alpha_i^h,$$

where $S(h, h) = s(h, h) = 1$ has been used. Using strict stationarity, this leads to

$$\begin{aligned}
\|\alpha_i \circ_{t,i} X_{t-i}\|_h^h &= \sum_{j=0}^{h-1} S(h, j) \sum_{l=0}^{j} s(j, l) \|X_t\|_l^l \alpha_i^j + \sum_{l=0}^{h-1} s(h, l) \|X_t\|_l^l \alpha_i^h + \alpha_i^h \|X_t\|_h^h \\
&\leq \sum_{j=0}^{h-1} \big|S(h, j)\big| \sum_{l=0}^{j} \big|s(j, l)\big| \|X_t\|_l^l \alpha_i^j + \sum_{l=0}^{h-1} \big|s(h, l)\big| \|X_t\|_l^l \alpha_i^h + \alpha_i^h \|X_t\|_h^h.
\end{aligned}$$

By using the above to bound the right-hand side of (B.7), isolating $\|X_t\|_h^h$ on one side of the inequality and using $(a+b)^{1/h} \leq a^{1/h} + b^{1/h}$ for $a, b \geq 0$ and $h \geq 1$, we get

$$\begin{aligned}
\|X_t\|_h \leq{} & \frac{1}{1 - \sum_{i=1}^{p} \alpha_i} \\
& \times \left(\sum_{i=1}^{p} \left(\sum_{j=0}^{h-1} \big|S(h, j)\big| \sum_{l=0}^{j} \big|s(j, l)\big| \|X_t\|_l^l \alpha_i^j + \sum_{l=0}^{h-1} \big|s(h, l)\big| \|X_t\|_l^l \alpha_i^h \right)^{1/h} + \|\epsilon_t\|_h \right).
\end{aligned}$$

As the last right-hand side above is composed of a finite weighted sum of $\|\epsilon_t\|_h$ and $\|X_t\|_j^j$, $j = 0, \dots, h-1$ only, we get $E(X_t^h) < \infty$ from $E(\epsilon_t^h) < \infty$ and using an induction-type argument.

Finally, to show the L_h-convergence, we use a similar argument. For notational convenience, we prove this assertion only for $p=2$, but the arguments transfer directly to $p>2$ as well. First, for all vectors $\boldsymbol{v}=(v_1,v_2)'\in\mathbb{R}^2$, we have

$$\begin{aligned}\left\|\boldsymbol{v}'\left(\boldsymbol{X}_t-\boldsymbol{X}_t^{(q)}\right)\right\|_h^h &= \left\|\boldsymbol{v}'[\mathbf{A}\circ]_{j=t-q}^{t}\boldsymbol{X}_{t-q+1}\right\|_h^h \\ &= \sum_{\boldsymbol{k}\in\mathbb{N}_0^p} E\left(\left(\boldsymbol{v}'\mathbf{A}\circ_t\boldsymbol{k}\right)^h\right)P\left([\mathbf{A}\circ]_{j=t-q}^{t-1}\boldsymbol{X}_{t-q+1}=\boldsymbol{k}\right).\end{aligned}\tag{B.9}$$

For the expectation on the last right-hand side, due to independence of the thinning operations, we can write

$$\begin{aligned}E\left(\left(\boldsymbol{v}'\mathbf{A}\circ_t\boldsymbol{k}\right)^h\right) &= E\left(\left(v_1(\alpha_1\circ_{t,1}k_1+\alpha_2\circ_{t,2}k_2)+v_2k_1\right)^h\right) \\ &= \sum_{r=0}^{h}\binom{h}{r}\sum_{u=0}^{r}\binom{r}{u}v_1^r(v_2k_1)^{h-r}E\left((\alpha_1\circ_{t,1}k_1)^u\right)E\left((\alpha_2\circ_{t,2}k_2)^{r-u}\right)\end{aligned}$$

and by using the same trick as above, i.e., the representation of moments of binomial distributions with the help of Stirling numbers, the last right-hand side can be expressed as

$$\sum_{r=0}^{h}\binom{h}{r}\sum_{u=0}^{r}\binom{r}{u}v_1^r(v_2k_1)^{h-r}\left(\sum_{a=0}^{u}S(u,a)\sum_{b=0}^{a}s(a,b)k_1^b\alpha_1^a\right)\left(\sum_{c=0}^{r-u}S(r-u,c)\sum_{d=0}^{c}s(c,d)k_2^d\alpha_2^c\right).$$

By extracting only the summand with $a=b=u$ and $c=d=r-u$, we get

$$\begin{aligned}&\sum_{r=0}^{h}\binom{h}{r}\sum_{u=0}^{r}\binom{r}{u}v_1^r(v_2)^{h-r} \\ &\quad\times\left(\sum_{\substack{a=0\ b=0\ c=0\ d=0 \\ (a,b,c,d)\neq(u,u,r-u,r-u)}}^{u\quad a\quad r-u\quad c} S(u,a)s(a,b)S(r-u,c)s(c,d)k_2^dk_1^{b+h-r}\alpha_1^a\alpha_2^c\right) \\ &\quad+\sum_{r=0}^{h}\binom{h}{r}\sum_{u=0}^{r}\binom{r}{u}v_1^r(v_2k_1)^{h-r}(k_1\alpha_1)^u(k_2\alpha_2)^{r-u} \\ &=\sum_{r=0}^{h}\binom{h}{r}\sum_{u=0}^{r}\binom{r}{u}v_1^r(v_2)^{h-r} \\ &\quad\times\left(\sum_{\substack{a=0\ b=0\ c=0\ d=0 \\ (a,b,c,d)\neq(u,u,r-u,r-u)}}^{u\quad a\quad r-u\quad c} S(u,a)s(a,b)S(r-u,c)s(c,d)k_2^dk_1^{b+h-r}\alpha_1^a\alpha_2^c\right) \\ &\quad+\left(\boldsymbol{v}'\mathbf{A}\boldsymbol{k}\right)^h.\end{aligned}$$

Together with (B.9) and Hölder inequality with $p = (b+h-r+d)/(b+h-r)$ and $q = (b+h-r+d)/d$ such that $1/p + 1/q = 1$, this leads to

$$
\begin{aligned}
&\left\| \boldsymbol{v}'[\mathbf{A}\circ]_{j=t-q}^{t} \boldsymbol{X}_{t-q+1} \right\|_h^h \\
&\quad = \sum_{r=0}^{h} \binom{h}{r} \sum_{u=0}^{r} \binom{r}{u} v_1^r (v_2)^{h-r} \\
&\qquad \times \sum_{\substack{a=0 \\ (a,b,c,d)\neq(u,u,r-u,r-u)}}^{u} \sum_{b=0}^{a} \sum_{c=0}^{r-u} \sum_{d=0}^{c} \left|S(r-u,c)\right|\left|s(c,d)\right|\left|S(u,a)\right|\left|s(a,b)\right| \alpha_1^a \alpha_2^c \\
&\qquad \times E\Big(\big((1,0)[\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1}\big)^{b+h-r} \big((0,1)[\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1}\big)^{d}\Big) \\
&\qquad + E\Big(\big(\boldsymbol{v}'\mathbf{A}[\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1}\big)^{h}\Big) \\
&\quad \leq \sum_{r=0}^{h} \binom{h}{r} \sum_{u=0}^{r} \binom{r}{u} v_1^r (v_2)^{h-r} \\
&\qquad \times \sum_{\substack{a=0 \\ (a,b,c,d)\neq(u,u,r-u,r-u)}}^{u} \sum_{b=0}^{a} \sum_{c=0}^{r-u} \sum_{d=0}^{c} \left|S(r-u,c)\right|\left|s(c,d)\right|\left|S(u,a)\right|\left|s(a,b)\right| \alpha_1^a \alpha_2^c \\
&\qquad \times \left\|(1,0)[\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1}\right\|_{b+h-r+d}^{b+h-r} \left\|(0,1)[\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1}\right\|_{b+h-r+d}^{d} \\
&\qquad + \left\|\boldsymbol{v}'\mathbf{A}[\mathbf{A}\circ]_{j=t-q}^{t-1} \boldsymbol{X}_{t-q+1}\right\|_h^h.
\end{aligned}
$$

As the latter inequality holds for arbitrary vector $\boldsymbol{v} \in \mathbb{R}^2$, we can use it successively to get

$$
\begin{aligned}
&\left\| \boldsymbol{u}'[\mathbf{A}\circ]_{j=t-q}^{t} \boldsymbol{X}_{t-q+1} \right\|_h^h \\
&\quad \leq \sum_{r=0}^{h} \binom{h}{r} \sum_{u=0}^{r} \binom{r}{u} \sum_{\substack{a=0 \\ (a,b,c,d)\neq(u,u,r-u,r-u)}}^{u} \sum_{b=0}^{a} \sum_{c=0}^{r-u} \sum_{d=0}^{c} \left|S(r-u,c)\right|\left|s(c,d)\right|\left|S(u,a)\right|\left|s(a,b)\right| \alpha_1^a \alpha_2^c \\
&\qquad \times \sum_{k=0}^{q} v_1(k)^r v_2(k)^{h-r} \left\|(1,0)[\mathbf{A}\circ]_{j=t-q}^{t-k-1} \boldsymbol{X}_{t-q+1}\right\|_{b+h-r+d}^{b+h-r} \\
&\qquad \times \left\|(0,1)[\mathbf{A}\circ]_{j=t-q}^{t-k-1} \boldsymbol{X}_{t-q+1}\right\|_{b+h-r+d}^{d} \\
&\qquad + \left\|\boldsymbol{u}'\mathbf{A}^{q+1} \boldsymbol{X}_{t-q+1}\right\|_h^h,
\end{aligned}
$$

where $v_1(k) = \boldsymbol{u}'\mathbf{A}^k(1,0)'$ and $v_2(k) = \boldsymbol{u}'\mathbf{A}^k(0,1)'$. By using $(a+b)^{1/h} \leq a^{1/h} + b^{1/h}$ for $a, b \geq 0$ and $h \geq 1$, we have

$$
\begin{aligned}
&\big\|\boldsymbol{u}'[\mathbf{A}\circ]_{j=t-q}^{t}\boldsymbol{X}_{t-q+1}\big\|_h \\
&\quad\leq \sum_{r=0}^{h}\sum_{u=0}^{r}\left(\binom{h}{r}\binom{r}{u}\right)^{1/h} \\
&\quad\times \sum_{\substack{a=0 \\ a\neq u, b\neq u}}^{u}\sum_{b=0}^{a}\sum_{\substack{c=0 \\ c\neq r-u, d\neq r-u}}^{r-u}\sum_{d=0}^{c}\big(\big|S(r-u,c)\big|\big|s(c,d)\big|\big|S(u,a)\big|\big|s(a,b)\big|\big)^{1/h}\alpha_1^{a/h}\alpha_2^{c/h} \\
&\quad\times \sum_{k=0}^{q} v_1(k)^{r/h}v_2(k)^{(h-r)/h}\big\|(1,0)[\mathbf{A}\circ]_{j=t-q}^{t-k-1}\boldsymbol{X}_{t-q+1}\big\|_{b+h-r+d}^{(b+h-r)/h} \\
&\quad\times \big\|(0,1)[\mathbf{A}\circ]_{j=t-q}^{t-k-1}\boldsymbol{X}_{t-q+1}\big\|_{b+h-r+d}^{d/h} \\
&\quad+ \big\|\boldsymbol{u}'\mathbf{A}^{q+1}\boldsymbol{X}_{t-q+1}\big\|_h.
\end{aligned}
$$

By Minkowski's inequality and stationarity, $\|\boldsymbol{u}'\mathbf{A}^{q+1}\boldsymbol{X}_{t-q+1}\|_h$ can be bounded by

$$\mathbf{1}'\mathbf{A}^{q+1}\mathbf{1}\|X_t\|_h \to 0,$$

as $q \to \infty$, because all the moduli of all eigenvalues of $\mathbf{A}$ are less than one. Further, by the same reasoning, we get $v_1(k), v_2(k) \leq \mathbf{1}'\mathbf{A}^k\mathbf{1}$ such that $v_1(k)^{r/h}v_2(k)^{(h-r)/h} \leq \mathbf{1}'\mathbf{A}^k\mathbf{1}$ as well. Hence, $v_1(k)^{r/h}v_2(k)^{(h-r)/h} \sim \theta^k$ with $\theta \in (0,1)$ is geometrically decreasing as k becomes large. Furthermore, due to extracting the summand with $a = b = u$ and $c = d = r - u$ from the sums in the first term on the last right-hand side above, $1 \leq b + h - r + d \leq h - 1$ does always hold. Altogether, by using an induction-type argument, this first term above converges to zero as $q \to \infty$ as it contains only moments up to order $h - 1$. This proves L_h-convergence. □

Lemma B.2 (Basic properties of the bootstrap INAR(p) **process).** *Under Assumption 2 with $s = 1$ and Assumption 3 with $s = 1$, the* INAR(p) *bootstrap recursion (3.1) has (with P-probability tending to one) a strictly stationary solution $(X_t^*, t \in \mathbb{Z})$ with $E^*(X_t^*) = \mu_{\epsilon^*}/(1 - \sum_{i=1}^{p}\widehat{\alpha}_i) = O_P(1)$. Precisely, the bootstrap* INAR(p) *process can be represented as an* INMA(∞) *process, that is,*

$$X_t^* = \sum_{k=0}^{\infty}\boldsymbol{u}'\big([\widehat{\mathbf{A}}\circ^*]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}^*, \qquad t \in \mathbb{Z}, \tag{B.10}$$

where $X_t^* = \boldsymbol{u}'\boldsymbol{X}_t^*$ with $\boldsymbol{X}_t^*$, $\boldsymbol{\epsilon}_t^*$, $\widehat{\mathbf{A}}$ and $[\widehat{\mathbf{A}}\circ^*]_{j=s}^{t}$ defined similarly as in Lemma B.1, but based on bootstrap quantities X_t^*, ϵ_t^*, estimators $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ and entry-wise bootstrap binomial thinning operations $\circ_t^*$. The convergence of (B.10) is in L_1-sense and almost surely in P-probability, respectively. If Assumption 3 holds for some $h \in \mathbb{N}$, i.e., $E^*(\epsilon_t^{*h}) = O_P(1)$, it holds $E^*(X_t^{*h}) = O_P(1)$ and the convergence of (B.10) holds in L_h-sense in P-probability.

Proof. Similar to the proof of Lemma B.1, we get

$$E^*\big(\big|X_t^* - X_t^{*(q)}\big|\big) \leq E^*\big(\big|\boldsymbol{X}_t^* - \boldsymbol{X}_t^{*(q)}\big|_1\big) = \mu_{X^*}\mathbf{1}_p'\widehat{\mathbf{A}}^{q+1}\mathbf{1}_p, \tag{B.11}$$

where $\mu_{X^*} = E^*(X_{t-q+1}^*) = O_P(1)$. As with P-probability tending to one, all eigenvalues of $\widehat{\mathbf{A}}$ have modulus less than one by Assumption 2 and Assumption 3(i), we get

$$\mu_{X^*}\mathbf{1}_p'\widehat{\mathbf{A}}^{q+1}\mathbf{1}_p = o_P(1),$$

which proves L_1-convergence in probability of (B.10). By similar arguments, we get

$$E^*\big(X_t^*\big) = \boldsymbol{u}'(\mathbf{I}_{p\times p} - \widehat{\mathbf{A}})^{-1}\boldsymbol{u}\mu_{\epsilon^*} = \frac{\mu_\epsilon^*}{1 - \sum_{i=1}^p \widehat{\alpha}_i}.$$

Further, from Assumption 3(i), we get that for all $\varepsilon > 0$, there exists a sequence $(\delta_n, n \in \mathbb{N})$ with $\delta_n \searrow 0$ such that $P_n(|\widehat{\mathbf{A}}_n - \mathbf{A}| < \delta_n) > 1 - \varepsilon$ for all $n \in \mathbb{N}$. By denoting $\Omega_{\varepsilon,n} = \{\omega : |\widehat{\mathbf{A}}_n - \mathbf{A}| < \delta_n\}$, we get by the same arguments used in the proof of Lemma B.1 and from Assumption 3(i) that for all sequences $(\omega_n, n \in \mathbb{N})$ with $\omega_n \in \Omega_{\varepsilon,n}$,

$$\sum_{k=0}^{\infty} \boldsymbol{u}'\big(\big[\widehat{\mathbf{A}}_n(\omega_n)\circ^*\big]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}^*$$

converges almost surely with respect to P^*. This proves almost sure convergence in probability of (B.10). To show that (B.10) solves the INAR(p) bootstrap recursion (3.1) is completely analogous to the calculations in (B.6). □

Lemma B.3 (INAR bootstrap: Weak convergence). *Suppose that either the assumptions of Corollary 3.4 or of Corollary 3.7 are satisfied. Then, we have $X_t^* \xrightarrow{d} X_t$ in probability as $n \to \infty$, i.e., it holds $F^*(x) := P^*(X_t^* \leq x) \xrightarrow{P} F(x) := P(X_t \leq x)$ as $n \to \infty$ for all continuity points x of $F(\cdot)$. Furthermore, for all $t_1, \dots, t_d \in \mathbb{Z}$, we have*

$$\big(X_{t_1}^*, \dots, X_{t_d}^*\big)' \xrightarrow{d} (X_{t_1}, \dots, X_{t_d})'$$

in probability.

Proof. Let x be a continuity point of $F(x) = P(X_t \leq x)$. Then, we have to show that for all $\varepsilon, \delta > 0$ there is an $n_0 = n_0(\delta, \varepsilon)$ such that

$$P\big(\big|P^*\big(X_t^* \leq x\big) - P(X_t \leq x)\big| \geq \varepsilon\big) < \delta \qquad \text{for all sample sizes } n > n_0.$$

Let $\varepsilon, \delta > 0$. From Lemma B.2, we get that the bootstrap INAR(p) process has an INMA(∞) representation, which we can decompose for any $q \in \mathbb{N}_0$ such that

$$\begin{aligned} X_t^* &= \sum_{k=0}^{q} \boldsymbol{u}'\big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}^* + \sum_{k=q+1}^{\infty} \boldsymbol{u}'\big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}^* \\ &= X_t^{*(q)} + \big(X_t^* - X_t^{*(q)}\big). \end{aligned} \tag{B.12}$$

Then, as $X_t^* \geq X_t^{*(q)}$, we can write

$$P^*\big(X_t^* \leq x\big) = P^*\big(X_t^{*(q)} \leq x\big) - P^*\big(x - \big(X_t^* - X_t^{*(q)}\big) < X_t^{*(q)} \leq x\big)$$

with $P^*(x - (X_t^* - X_t^{*(q)}) < X_t^{*(q)} \leq x) \leq P^*((X_t^* - X_t^{*(q)}) \geq 1) \leq E^*(X_t^* - X_t^{*(q)})$. Further, similar to (B.5), we get

$$\begin{aligned} E^*\big(\boldsymbol{X}_t^* - \boldsymbol{X}_t^{*(q)}\big) &= E^*\big(\big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-q}^{t}\big)\boldsymbol{\epsilon}_{t-q-1}^*\big) + E^*\left(\sum_{k=q+2}^{\infty} \big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}^*\right) \\ &= E^*\big(\big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-q}^{t}\big)\boldsymbol{\epsilon}_{t-q-1}^*\big) + \widehat{\mathbf{A}}E^*\left(\sum_{k=q+1}^{\infty} \big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\big)\boldsymbol{\epsilon}_{t-k}^*\right) \\ &= E^*\big(\big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-q}^{t}\big)\boldsymbol{\epsilon}_{t-q-1}^*\big) + \widehat{\mathbf{A}}E^*\big(\boldsymbol{X}_t^* - \boldsymbol{X}_t^{*(q)}\big) \end{aligned}$$

leading to

$$\begin{aligned} E^*\big(X_t^* - X_t^{*(q)}\big) &= \boldsymbol{u}'E^*\big(\boldsymbol{X}_t^* - \boldsymbol{X}_t^{*(q)}\big) = \boldsymbol{u}'(\mathbf{I} - \widehat{\mathbf{A}})^{-1}E^*\big(\big(\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-q}^{t}\big)\boldsymbol{\epsilon}_{t-q-1}^*\big) \\ &= \mu_\epsilon^*\boldsymbol{u}'(\mathbf{I} - \widehat{\mathbf{A}})^{-1}\widehat{\mathbf{A}}^{q+1}\boldsymbol{u}. \end{aligned}$$

Analogously, we have

$$P(X_t \leq x) = P\big(X_t^{(q)} \leq x\big) - P\big(x - \big(X_t - X_t^{(q)}\big) < X_t^{(q)} \leq x\big)$$

with $P(x - (X_t - X_t^{(q)}) < X_t^{(q)} \leq x) \leq E(X_t - X_t^{(q)})$ and $E(X_t - X_t^{(q)}) = \mu_\epsilon\boldsymbol{u}'(\mathbf{I} - \mathbf{A})^{-1}\mathbf{A}^{q+1}\boldsymbol{u}$. Altogether, this leads to

$$\begin{aligned} &P\big(\big|P^*\big(X_t^* \leq x\big) - P(X_t \leq x)\big| \geq \varepsilon\big) \\ &\quad\leq P\left(\big|P^*\big(X_t^{*(q)} \leq x\big) - P\big(X_t^{(q)} \leq x\big)\big| \geq \frac{\varepsilon}{3}\right) + P\left(\big|\mu_\epsilon^*\boldsymbol{u}'(\mathbf{I} - \widehat{\mathbf{A}})^{-1}\widehat{\mathbf{A}}^{q+1}\boldsymbol{u}\big| \geq \frac{\varepsilon}{3}\right) \\ &\qquad + P\left(\big|\mu_\epsilon\boldsymbol{u}'(\mathbf{I} - \mathbf{A})^{-1}\mathbf{A}^{q+1}\boldsymbol{u}\big| \geq \frac{\varepsilon}{3}\right) \\ &\quad=: \mathrm{I} + \mathrm{II} + \mathrm{III}. \end{aligned}$$

Now, note that $\widehat{\mathbf{A}} \xrightarrow{P} \mathbf{A}$ and $\mu_\epsilon^* \to \mu_\epsilon < \infty$ hold in the parametric setup of Corollary 3.4 and in the semi-parametric setup of Corollary 3.7. On the one hand, as all eigenvalues of $\mathbf{A}$ have modulus less than one, we can choose $\kappa > 0$ sufficiently small such that all $(p \times p)$ matrices $\mathbf{B}$ with $\|\mathbf{B} - \mathbf{A}\| \leq \kappa$ have all eigenvalues less than and uniformly away from one. In particular, this assures existence of $(\mathbf{I} - \widehat{\mathbf{A}})^{-1}$ in the above with probability tending to one. On the other hand, we can choose $M < \infty$ such that $P(\mu_\epsilon^* > M) < \delta$. Considering term II first, for sufficiently large $n = n(\kappa, M)$ and sufficiently large $q = q(\varepsilon, \delta, \kappa, M)$, this leads to

$$\begin{aligned}
\mathrm{II} &\leq P\left(\left|\mu_\epsilon^* \boldsymbol{u}'(\mathbf{I} - \widehat{\mathbf{A}})^{-1}\widehat{\mathbf{A}}^{q+1}\boldsymbol{u}\right| \geq \frac{\varepsilon}{3}, \|\widehat{\mathbf{A}} - \mathbf{A}\| \leq \kappa, \mu_\epsilon^* \leq M\right) \\
&\quad + P(\|\widehat{\mathbf{A}} - \mathbf{A}\| > \kappa) + P\big(\mu_\epsilon^* > M\big) \\
&< P\left(\left|\mu_\epsilon^* \boldsymbol{u}'(\mathbf{I} - \widehat{\mathbf{A}})^{-1}\widehat{\mathbf{A}}^{q+1}\boldsymbol{u}\right| \geq \frac{\varepsilon}{3}, \|\widehat{\mathbf{A}} - \mathbf{A}\| \leq \kappa, \mu_\epsilon^* \leq M\right) + 2\delta \\
&< 3\delta.
\end{aligned}$$

As $\mu_\epsilon \boldsymbol{u}'(\mathbf{I} - \mathbf{A})^{-1}\mathbf{A}^{q+1}\boldsymbol{u}$ is deterministic, we can find also q large enough such that $|\mu_\epsilon \boldsymbol{u}'(\mathbf{I} - \mathbf{A})^{-1}\mathbf{A}^{q+1}\boldsymbol{u}| < \varepsilon/3$ holds leading to $\mathrm{III} = 0$. Hence, it remains to treat I. As $X_t^{*(q)}$ and $X_t^{(q)}$ are both finite sums of independent summands (see, e.g., (B.12)), by Cramér–Wold device, it remains to show that for all fixed $k \in \mathbb{N}_0$ that

$$\boldsymbol{u}'\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1}^* \xrightarrow{d} \boldsymbol{u}'[\mathbf{A}\circ]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1}$$

in probability, to complete the proof. That is, we have to show that for all continuity points of $P(\boldsymbol{u}'[\mathbf{A}\circ]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1} \leq \cdot)$, we have

$$P^*\big(\boldsymbol{u}'\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1}^* \leq x\big) \xrightarrow{P} P\big(\boldsymbol{u}'[\mathbf{A}\circ]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1} \leq x\big).$$

Note here that the continuity points of $P(X_t \leq \cdot)$ and $P(\boldsymbol{u}'[\mathbf{A}\circ]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1} \leq \cdot)$, $k \in \mathbb{N}_0$ simply coincide as all corresponding random variables have range $\mathbb{N}_0$. Conditioning on ϵ_{t-k+1}^* (recall that $\boldsymbol{\epsilon}_{t-k+1}^* = (\epsilon_{t-k+1}^*, 0, \ldots, 0)'$) and independence of thinning operations and innovations leads to

$$P^*\big(\boldsymbol{u}'\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1}^* \leq x\big) = \sum_{s \in \mathbb{N}_0} P^*\big(\boldsymbol{u}'\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}(s\boldsymbol{u}) \leq x\big) P^*\big(\epsilon_{t-k+1}^* = s\big)$$

with an analogous representation also for $P(\boldsymbol{u}'[\mathbf{A}\circ]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1} \leq x)$. Now, using both representations, for any $K \in \mathbb{N}_0$, we get

$$\begin{aligned}
&\big|P^*\big(\boldsymbol{u}'\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1}^* \leq x\big) - P\big(\boldsymbol{u}'[\mathbf{A}\circ]_{j=t-k+1}^{t}\boldsymbol{\epsilon}_{t-k+1} \leq x\big)\big| \\
&\quad \leq \sum_{s \in \mathbb{N}_0} P^*\big(\boldsymbol{u}'\big[\widehat{\mathbf{A}}\circ^*\big]_{j=t-k+1}^{t}(s\boldsymbol{u}) \leq x\big)\big|P^*\big(\epsilon_{t-k+1}^* = s\big) - P(\epsilon_{t-k+1} = s)\big|
\end{aligned}$$

$$+ \sum_{\substack{s\in\mathbb{N}_0\\ s\le K}} \left|P^*\left(\boldsymbol{u}'\left[\widehat{\mathbf{A}}\circ^*\right]_{j=t-k+1}^{t}(s\boldsymbol{u}) \le x\right) - P\left(\boldsymbol{u}'[\mathbf{A}\circ]_{j=t-k+1}^{t}(s\boldsymbol{u}) \le x\right)\right| P(\epsilon_{t-k+1} = s)$$

$$+ \sum_{\substack{s\in\mathbb{N}_0\\ s> K}} \left|P^*\left(\boldsymbol{u}'\left[\widehat{\mathbf{A}}\circ^*\right]_{j=t-k+1}^{t}(s\boldsymbol{u}) \le x\right) - P\left(\boldsymbol{u}'[\mathbf{A}\circ]_{j=t-k+1}^{t}(s\boldsymbol{u}) \le x\right)\right| P(\epsilon_{t-k+1} = s).$$

Now, for fixed $\delta > 0$, the aim is to show that we can choose $K = K(\delta)$ and $n = n(K, \delta)$ large enough such that the last right-hand side above is smaller than δ. The first term above can be bounded by $\sum_{s\in\mathbb{N}_0} |P^*(\epsilon^*_{t-k+1} = s) - P(\epsilon_{t-k+1} = s)|$, which equals $\sum_{s\in\mathbb{N}_0} |G_{\widehat{\boldsymbol{\theta}}}(s) - G_{\boldsymbol{\theta}_0}(s)| = o_P(1)$ in the parametric setup and $\sum_{s\in\mathbb{N}_0} |\widehat{G}_{\mathrm{sp}}(s) - G(s)| = o_P(1)$ in the semi-parametric setup. The third term can be bounded by $\sum_{s\in\mathbb{N}_0, s>K} P(\epsilon_{t-k+1} = s)$ which becomes arbitrary small for K chosen sufficiently large in the parametric and semi-parametric setup. For all fixed $K \in \mathbb{N}_0$, the second term above is a finite sum and it remains to show that

$$P^*\left(\boldsymbol{u}'\left[\widehat{\mathbf{A}}\circ^*\right]_{j=t-k+1}^{t}(s\boldsymbol{u}) \le x\right) \xrightarrow{P} P\left(\boldsymbol{u}'[\mathbf{A}\circ]_{j=t-k+1}^{t}(s\boldsymbol{u}) \le x\right)$$

for all fixed $k \in \mathbb{N}$ and $s \in \mathbb{N}_0$. The latter can be shown by successively conditioning and decomposing using similar arguments as above.

Finally, the joint convergence can be shown by the Cramér–Wold device and by following the lines of the proof above. □

Appendix C: Constructing INAR residuals: Some (naïve) ideas

In view of the popularity of the AR bootstrap, it would be desirable to adapt it to the INAR case. For this purpose, it would be necessary (among others) to construct an appropriate type of residuals from INAR time series data. In the sequel, we shall briefly discuss three such approaches, which appear more or less obvious at a first glance. However, it turns out that all these approaches suffer from severe handicaps, so a straightforward adaption of a model-based AR bootstrap to the set-up of INAR processes is not possible.

C.1. Exploiting discreteness

The randomness of the thinning operation in INAR models does generally not allow to estimate consistently the innovations ϵ_t. However, it is possible to exploit the discreteness of INAR models to observe directly some of them even without any deviation. For instance, for INAR(1) models from (1.1), we have

$$X_t = \alpha \circ X_{t-1} + \epsilon_t = \epsilon_t \qquad \text{if } X_{t-1} = 0. \tag{C.1}$$

That means, whenever $X_{t-1} = 0$, we observe the true innovation without any deviation as $X_t = \epsilon_t$ one step ahead. As the range of ϵ_t is $\mathbb{N}_0$, the same holds for X_t leading to $p_0 = P(X_{t-1} = 0) > 0$ iff $P(\epsilon_t = 0) > 0$. That is, we can expect about $(n-1)p_0$ zeros in a sample $X_1, \ldots, X_{n-1}$, and this increases to ∞ as $n \to \infty$. This means, although it is not possible to estimate the innovations

ϵ_t consistently, it is still possible to observe asymptotically infinitely many of these innovations, which of course carry the true distribution. Hence, this approach would allow for asymptotically meaningful inference on the innovations' distribution. In the context of random coefficient AR models, a similar approach has been employed by Fink and Kreiß [16] for bootstrapping of random coefficient AR models. However, in finite samples, this approach suffers from severe drawbacks:

(i) If p_0 is small, only few innovations are observed in an INAR sample $X_1, \dots, X_n$. Consequently, the effective sample size for the sequence of innovations is very small, leading to inferior finite sample performance.
(ii) In principle, the idea of (C.1) can be extended to INAR(p) models of general order p. In this case, to be able to observe the innovations ϵ_t directly, one even needs $X_{t-1} = \cdots = X_{t-p} = 0$. But the latter event is by far rarer than $X_{t-1} = 0$.
(iii) In general, this approach suffers from a considerable loss in efficiency as only few information from a data sample $X_1, \dots, X_n$ is eventually used.

C.2. AR fitting and rounding

Motivated by the fact that INAR(p) and AR(p) models share the same autocorrelation function, Cardinal, Roy and Lambert [9], Kim and Park [28,29] proposed to fit an AR(p) model (1.5) to the INAR sample $X_1, \dots, X_n$ to get quantities $\widehat{e}_t + \widehat{\nu} = X_t - \widehat{\alpha}_1 X_{t-1} - \cdots - \widehat{\alpha}_p X_{t-p}$; see also Park and Kim [42] for the notion of 'expected residuals'. A sequence of residuals having range $\mathbb{N}_0$ is constructed by

$$\breve{\epsilon}_t := \begin{cases} [\widehat{e}_t + \widehat{\nu}], & \widehat{e}_t + \widehat{\nu} > 0, \\ 0, & \widehat{e}_t + \widehat{\nu} \leq 0, \end{cases}$$

where $[x]$ denotes a rounding operator to ensure that $\breve{\epsilon}_t$ is integer-valued. In their papers, Cardinal, Roy and Lambert [9], Kim and Park [28] used this sequence of residuals for model-based bootstrapping, whereas Kim and Park [29] adapt this approach for bootstrapping signed INAR(p) models that have range $\mathbb{Z}$. But as illustrated in Figure 1(a), this method for constructing residuals will generally not work sufficiently precise. For the example presented in Figure 1(a), it was assumed that $\epsilon_t \sim \text{Poi}(1.0)$, and the corresponding true probability mass function (PMF) of the innovations is plotted as thick light bars. Now assume an INAR(1) model with either $\alpha = 0.4$ or $\alpha = 0.8$, then the distributions of $[X_t - \alpha X_{t-1}]$ (plotted with thin bars in black or dark gray, respectively) strongly deviate from the one of ϵ_t. Moreover, it can be easily seen that no simple transformation as, e.g., shifting the distribution or setting negative values to zero does cure the problem.

C.3. Additional thinning

Another naïve approach that tries to mimic the randomness of the thinning operation in the construction of residuals is as follows. Fit an INAR(p) model (1.3) to the sample $X_1, \dots, X_n$ to

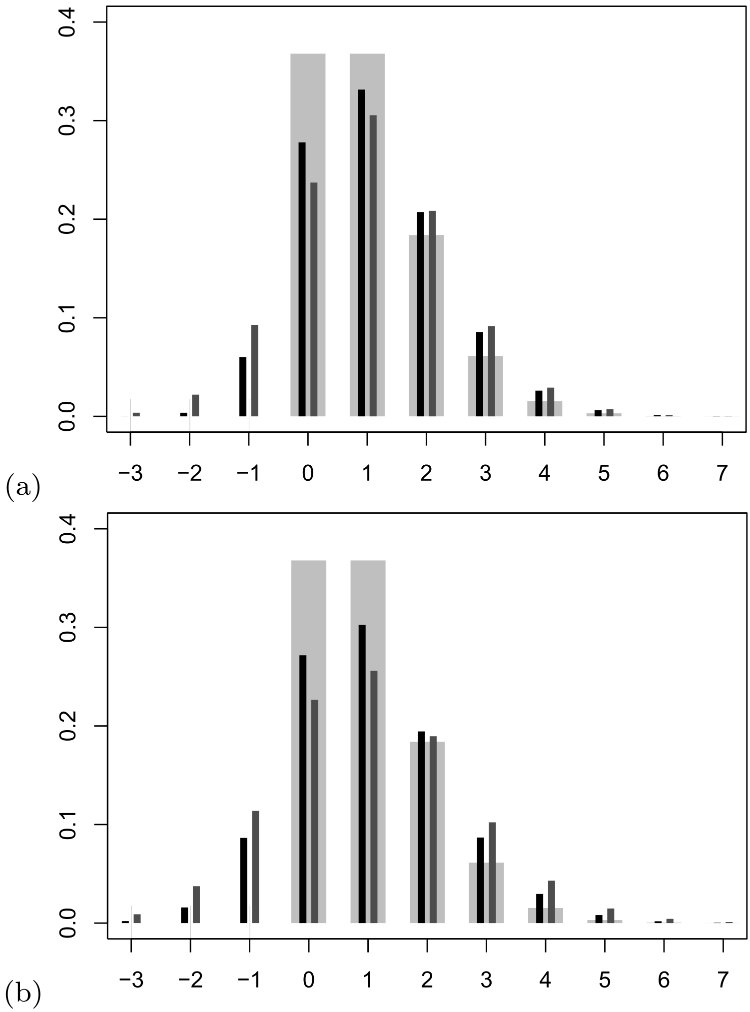

Figure 1. PMF of $\epsilon_t \sim \mathrm{Poi}(1.0)$ (light gray) against PMF of rounded residuals $[X_t - \alpha X_{t-1}]$ in (a), and thinning-based residuals $X_t - \alpha \,\widetilde{\circ}\, X_{t-1}$ in (b); here $\alpha = 0.4$ (black) or $\alpha = 0.8$ (dark gray).

get estimated coefficients $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$. Then, define a sequence of residuals by

$$\widetilde{\epsilon}_t := X_t - \widehat{\alpha}_1 \,\widetilde{\circ}\, X_{t-1} - \cdots - \widehat{\alpha}_p \,\widetilde{\circ}\, X_{t-p}, \qquad t = p+1, \ldots, n, \tag{C.2}$$

where "$\widetilde{\circ}$" denotes mutually independent binomial thinning operations executed independently of everything else; see Bisaglia and Gerolimetto [4] for such an approach. But again, this approach is not applicable, because, among others, the residuals $\widetilde{\epsilon}_{1+p}, \ldots, \widetilde{\epsilon}_n$ become negative with positive (asymptotically non-vanishing) probability. Setting the negative values to zero does not cure this problem, also see the example plotted in Figure 1(b), which plots the PMF of $X_t - \alpha \widetilde{\circ} X_{t-1}$ for $\alpha = 0.4$ (black) or $\alpha = 0.8$ (dark gray). This immediately leads to the conclusion that the INAR residuals in (C.2) do not mimic properly the true marginal distribution of the innovations, such that these residuals are not appropriate for statistical inference on the innovations' distribution. Additionally, residuals from (C.2) are random and hence not reproducible.

Appendix D: About AR bootstrap (in)consistency for INAR time series data

In the following lemma, we show that an AR bootstrap applied to INAR data (see Section 2 for background information) is generally sufficient to lead to valid bootstrap approximations for the statistic of the sample mean.

Lemma D.1 (AR bootstrap consistency for the sample mean). *Suppose* $(X_t, t \in \mathbb{Z})$ *is a stationary* INAR(p) *model as in* (1.3) *with* $E(\epsilon_t^4) < \infty$. *Then, we have* $n \operatorname{Var}^+(\overline{Y}^+) \to V$ *and*

$$\sqrt{n}\overline{Y}^+ = \sqrt{n}\big(\overline{X}^+ - \overline{X}\big) \xrightarrow{d} \mathcal{N}(0, V)$$

in probability, respectively, leading to

$$d_K\big(\mathcal{L}^+\big(\sqrt{n}\overline{Y}^+\big), \mathcal{L}\big(\sqrt{n}(\overline{X} - \mu_X)\big)\big) := \sup_{x \in \mathbb{R}} \big| P^+\big(\sqrt{n}\overline{Y}^+ \le x\big) - P\big(\sqrt{n}(\overline{X} - \mu_X) \le x\big)\big| = o_P(1),$$

where d_K *denotes the Kolmogorov-Smirnov distance and* $\mathcal{L}(X)$ *denotes the probability law of a random variable* X. *Here and throughout the paper,* $\mathcal{L}^+$, P^+, E^+, Var^+, *etc. denote bootstrap law, bootstrap probability measure, bootstrap mean, bootstrap variance, etc., respectively, induced by the AR bootstrap* (*conditional on the sample* $X_1, \dots, X_n$).

Proof. Similar to the proof of Theorem 3.1 in Kreiß, Paparoditis and Politis [34], we get that

$$\sqrt{n}\overline{Y}^+ \xrightarrow{d} \mathcal{N}(0, \widetilde{V}), \qquad \text{in prob.},$$

where $\widetilde{V} = \sum_{h=-\infty}^{\infty} \gamma_{\widetilde{Y}}(h)$ with $\gamma_{\widetilde{Y}}(h) = \operatorname{Cov}(\widetilde{Y}_{t+h}, \widetilde{Y}_t)$. Here, $(\widetilde{Y}_t, t \in \mathbb{Z})$ denotes the *companion process* of $(Y_t^+, t \in \mathbb{Z})$ in the sense of Kreiß, Paparoditis and Politis [34]. That is,

$$\widetilde{Y}_t = \alpha_1 \widetilde{Y}_{t-1} + \cdots + \alpha_p \widetilde{Y}_{t-p} + \widetilde{u}_t, \qquad t \in \mathbb{Z}, \tag{D.1}$$

where $(\widetilde{u}_t)$ consists of i.i.d. random variables whose *marginal* distribution is identical to that of u_t with

$$\begin{aligned} u_t &:= X_t - \nu - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} \\ &= X_t - \mu_X - \alpha_1(X_{t-1} - \mu_X) - \cdots - \alpha_p(X_{t-p} - \mu_X). \end{aligned} \tag{D.2}$$

Now, observe that the INAR(p) process $(X_t, t \in \mathbb{Z})$ in (1.3) and its companion process $(\widetilde{Y}_t, t \in \mathbb{Z})$ in (D.1) share the same autocorrelation function, that is $\rho_X(h) = \rho_{\widetilde{Y}}(h)$, $h \in \mathbb{Z}$. Hence, it remains to show that $\gamma_X(0) = \gamma_{\widetilde{Y}}(0)$ holds to complete the proof. From the Yule–Walker equations and causality of the companion process in (D.1) assured by $\alpha_i \in (0, 1)$ with $\sum_{i=1}^{p} \alpha_i < 1$, we get

$$\gamma_{\widetilde{Y}}(0) = \alpha_1 \gamma_{\widetilde{Y}}(1) + \cdots + \alpha_p \gamma_{\widetilde{Y}}(p) + \sigma_{\widetilde{u}}^2, \tag{D.3}$$

where $\sigma_{\widetilde{u}}^2 = \mathrm{Var}(\widetilde{u}_t)$. Finally, exploiting $\mathcal{L}(\widetilde{u}_t) = \mathcal{L}(u_t)$, $E(u_t) = 0$, $\mathrm{Var}(u_t) \in (0, \infty)$ and (D.2), we get

$$\sigma_{\widetilde{u}}^2 = \mathrm{Var}(u_t) = \gamma_X(0) - \alpha_1 \gamma_X(1) - \cdots - \alpha_p \gamma_X(p)$$

leading together with (D.3) to $\gamma_{\widetilde{Y}}(0) = \gamma_X(0)$. □

While an AR bootstrap applied to INAR data is generally sufficient to lead to valid bootstrap approximations for the statistic of the sample mean, the situation becomes much different for statistics that depend also on other distributional features of the DGP beyond second-order structure. To illustrate this, we consider the sample variance $\widehat{\gamma}(0) = n^{-1} \sum_{t=1}^{n} (X_t - \overline{X})^2$ in the following example. For an INAR(1) model with Poisson-distributed innovations, it shows that an AR bootstrap cannot be generally valid for statistical inference for INAR data.

Example D.2 (AR bootstrap inconsistency for the sample variance). Suppose $(X_t, t \in \mathbb{Z})$ is a stationary Poisson INAR(1) process as in (1.1) with $\epsilon_t \sim \mathrm{Poi}(\lambda)$, $\lambda > 0$ and thinning parameter $\alpha \in (0, 1)$. This leads to $\mu_X = E(X_t) = \lambda/(1 - \alpha)$, and the true limiting variance of the sample variance $\widehat{\gamma}(0)$ is given by

$$n \,\mathrm{Var}\big(\widehat{\gamma}(0)\big) \to 2\mu_X^2 \frac{1 + \alpha^2}{1 - \alpha^2} + \mu_X \frac{1 + \alpha}{1 - \alpha}, \tag{D.4}$$

see Weiß and Schweer [54]. However, if an ordinary AR bootstrap (of order one) is applied to INAR(1) data, we get

$$n \,\mathrm{Var}^+\big(\widehat{\gamma}^+(0)\big) \to 2\mu_X^2 \frac{1 + \alpha^2}{1 - \alpha^2} + \mu_X \frac{1 - 4\alpha^2 + 6\alpha^3 - 3\alpha^4}{(1 - \alpha^2)^2} \tag{D.5}$$

in probability, see the proof below. A comparison of (D.4) and (D.5) shows that their respective second summands differ, leading to the conclusion that the AR bootstrap is not asymptotically valid. As an illustrative example, consider the case of $\lambda = 1$ and $\alpha \in \{0.4, 0.8\}$, leading to $\mu_X = 1.\overline{6}$ and $\mu_X = 5$, respectively. Then the asymptotic variances according to (D.4) and (D.5) are equal to about 11.56 vs. 9.25 ($\alpha = 0.4$) and 272.78 vs. 238.70 ($\alpha = 0.8$), respectively. Hence, the AR bootstrap systematically underestimates the variances in both cases. This difference does not only hold asymptotically, but also manifests itself for finite sample size. In Table 7, we summarize some results from the simulation study as described in Section 4.1. The columns "Simul. value" give the value of $n \,\mathrm{Var}(\widehat{\gamma}(0))$ that was empirically computed from the respective simulation runs, while the remaining columns give the median value from bootstrap estimates for $n \,\mathrm{Var}(\widehat{\gamma}(0))$. The columns "AR" refer

Table 7. Poisson INAR(1) model: Simulated values of $n\,\mathrm{Var}(\widehat{\gamma}(0))$ against the median of the corresponding bootstrap estimates, see Example D.2 for details

		$(\lambda,\alpha)=(1,0.4)$						$(\lambda,\alpha)=(1,0.8)$				
	Simul.	Median of bootstrap estimates					Simul.	Median of bootstrap estimates				
n	value	spINAR	pINAR	CBB	AR	MB	value	spINAR	pINAR	CBB	AR	MB
100	10.51	9.92	10.58	6.12	7.17	9.04	232.35	167.40	183.45	73.81	116.78	121.83
250	10.97	10.99	10.93	8.23	8.33	11.70	245.88	224.70	222.36	121.26	168.73	174.14
500	11.26	11.23	11.31	9.61	8.68	12.97	280.65	255.53	250.01	167.15	202.75	211.16
1000	11.44	11.36	11.50	9.82	9.00	12.55	277.70	266.38	262.94	191.45	222.16	236.72
∞	11.56				9.25		272.78				238.70	

to the AR bootstrap discussed here, and it becomes clear that the AR bootstrap estimates for $n\,\mathrm{Var}(\widehat{\gamma}(0))$ are much below the simulated true value also for finite n. The columns "CBB" and "MB" refer to the natural non-parametric bootstrap competitors of the circular block bootstrap and the Markov bootstrap, respectively (see Section 4.1 for more details), while "spINAR" and "pINAR" refer to the INAR bootstraps from Section 3. It is obvious that these methods produce estimates being much closer to the simulated values for $n\,\mathrm{Var}(\widehat{\gamma}(0))$.

Proof. For a (centered) linear process, where $X_t=\sum_{j=-\infty}^{\infty}\psi_j e_{t-j}$ with $(e_t, t\in\mathbb{Z})$ being i.i.d. with $E(e_t^4)<\infty$ and $\sum_{j=-\infty}^{\infty}|\psi_j|<\infty$, it is well known that

$$n\,\mathrm{Var}\big(\widehat{\gamma}(0)\big)\to 2\sum_{h=-\infty}^{\infty}\gamma^2(h)+\left(\frac{E(e_t^4)}{(E(e_t^2))^2}-3\right)\gamma(0)^2, \tag{D.6}$$

see Proposition 7.3.1 in Brockwell and Davis [5], for instance. For a causal AR(1) process, we have $\psi_j=\alpha^j$ for $j\geq 0$ and 0 otherwise, and it follows that $\gamma(h)=\alpha^h\gamma(0)$ with

$$\gamma(0)=E\big(e_t^2\big)\sum_{j=0}^{\infty}\psi_j^2=\frac{E(e_t^2)}{1-\alpha^2}.$$

So (D.6) becomes

$$n\,\mathrm{Var}\big(\widehat{\gamma}(0)\big)\to 2\gamma^2(0)\frac{1+\alpha^2}{1-\alpha^2}+\frac{E(e_t^4)-3(E(e_t^2))^2}{(1-\alpha^2)^2}. \tag{D.7}$$

Now we adapt (D.7) to the AR(1) bootstrap, as it is applied to the DGP Poisson INAR(1). With analogous arguments as in the proof of Theorem 3.1 in Kreiß, Paparoditis and Politis [34] as well as Lemma 5.3 in Bühlmann [7,8], we have $\mathrm{Var}^+(e_t^+)=E^+((e_t^+)^2)\xrightarrow{P}E(e_t^2)$ and

$E^+((e_t^+)^4) \xrightarrow{P} E(e_t^4)$, where $e_t = (X_t - \mu_X) - \alpha(X_{t-1} - \mu_X)$. So we can further evaluate $E(e_t^2)$ to obtain

$$E^+\big((e_t^+)^2\big) \to \big(1+\alpha^2\big)\gamma_X(0) - 2\alpha\gamma_X(1) = \big(1-\alpha^2\big)\gamma_X(0),$$

using that $\gamma_X(1) = \alpha\gamma_X(0)$. To compute $E^+((e_t^+)^4)$, let us introduce the notation

$$\bar{\mu}(s_1, \ldots, s_{r-1}) := E\big((X_t - \mu_X)(X_{t+s_1} - \mu_X)\cdots(X_{t+s_{r-1}} - \mu_X)\big)$$

for the joint central moments. Then we obtain

$$E^+\big((e_t^+)^4\big) \to \big(1+\alpha^4\big)\bar{\mu}(0,0,0) - 4\alpha\bar{\mu}(1,1,1) + 6\alpha^2\bar{\mu}(0,1,1) - 4\alpha^3\bar{\mu}(0,0,1).$$

According to Proposition 1 in Weiß [51], the Poisson INAR(1) process satisfies

$$\bar{\mu}(k,l,m) = \mu_X\cdot\alpha^m + \mu_X^2\cdot\alpha^{m-l+k} + 2\mu_X^2\cdot\alpha^{m+l-k} \qquad \text{for } 0\le k\le l\le m,$$

while $\gamma(0) = \mu_X$. So after some computations, we obtain the limit

$$E^+\big((e_t^+)^4\big) - 3\big(E^+\big((e_t^+)^2\big)\big)^2 \to \big(1 - 4\alpha^2 + 6\alpha^3 - 3\alpha^4\big)\mu_X.$$

Altogether, using (D.7), it follows that

$$n\,\mathrm{Var}^+\big(\widetilde{\gamma}^+(0)\big) \to 2\mu_X^2\frac{1+\alpha^2}{1-\alpha^2} + \mu_X\frac{1-4\alpha^2+6\alpha^3-3\alpha^4}{(1-\alpha^2)^2},$$

so the proof of (D.5) is complete. □

Appendix E: Further details about simulations and data example

E.1. Semi-parametric estimation procedure

Given a time series $x_1, \ldots, x_n$, the estimation procedure of Drost, van den Akker and Werker [13] computes a distribution with support in $\{0, \ldots, \max\{x_1, \ldots, x_n\}\}$ for the innovations, and values for the autoregressive parameters $\alpha_1, \ldots, \alpha_p$. As initial values for the α_i, Drost, van den Akker and Werker [13] suggest to use the corresponding Yule–Walker estimates $\widehat{\alpha}_{1;\mathrm{YW}}, \ldots, \widehat{\alpha}_{p;\mathrm{YW}}$ as obtained from the sample autocorrelation function. As the initial distribution, Drost, van den Akker and Werker [13] recommend the use of the uniform distribution on $\{0, \ldots, \max\{x_1, \ldots, x_n\}\}$, which, however, will usually deviate heavily from the true distribution of the ϵ. We experimented with two other non-parametric initializations:

- Poisson approximation: use the Poisson distribution with mean $\overline{x}(1-\widehat{\alpha}_{1;\mathrm{YW}}-\cdots-\widehat{\alpha}_{p;\mathrm{YW}})$, truncate it to $\{0,\ldots,\max\{x_1,\ldots,x_n\}\}$.
- ϵ-identification: Given the values of $x_t; x_{t-1},\ldots,x_{t-p}$ for $p+1\le t\le n$, consider the probability $f_t(i) := P(X_t = x_t \mid X_{t-1} = x_{t-1},\ldots,X_{t-p} = x_{t-p}, \epsilon_t = i)$ as a function of i (by using $\widehat{\alpha}_{1;\mathrm{YW}},\ldots,\widehat{\alpha}_{p;\mathrm{YW}}$ instead of the true autoregressive parameters), and define $\hat{\epsilon}_t := \arg\max_i f_t(i)$. Use the frequency distribution of $\hat{\epsilon}_{p+1},\ldots,\hat{\epsilon}_n$.

We compared both initialization approaches via simulations and realized nearly no difference in the ML estimates resulting from the approach by Drost, van den Akker and Werker [13]. Therefore, we finally used the second one for initialization. For autoregressive order $p=1$, we immediately obtain that

$$P(X_t = k \mid X_{t-1} = l_1, \epsilon_t = i) = \binom{l_1}{k-i}\alpha^{k-i}(1-\alpha)^{l-k+i} \tag{E.1}$$

for $\max\{0,k-l_1\}\le i\le k$, and 0 otherwise. Similarly, for $p=2$, we have

$$\begin{aligned} &P(X_t = k \mid X_{t-1} = l_1, X_{t-2} = l_2, \epsilon_t = i) \\ &= \sum_{r=\max\{0,k-i-l_2\}}^{\min\{l_1,k-i\}} \binom{l_1}{r}\alpha_1^r(1-\alpha_1)^{l_1-r}\cdot\binom{l_2}{k-i-r}\alpha_2^{k-i-r}(1-\alpha_2)^{l_2-k+i+r} \end{aligned} \tag{E.2}$$

for $\max\{0,k-l_1-l_2\}\le i\le k$, and 0 otherwise. Note that both distributions (E.1) and (E.2) might become one-point distributions (i.e., support of size 1), which implies that ϵ_t is identified exactly. This happens, for instance, if $l_1=\cdots=l_p=0$, i.e., if the last p observations were equal 0, also see Appendix C.1.

For numerical computation of the estimates, we used MATLAB's `fmincon` function (constrained nonlinear optimization, algorithm `active-set`), where the constraints were chosen such that all parameter restrictions of a stationary INAR(p) model are satisfied (the boxes for each estimate were bounded by 10^{-6} and $1-10^{-6}$). Remember that the coefficients $\alpha_1,\ldots,\alpha_p>0$ of an INAR(p) model have to satisfy the condition $\sum_{j=1}^p\alpha_j<1$ to be strictly stationary.

E.2. Centering schemes

In the sequel, we state briefly the different centering schemes for the semi-parametric INAR-Bootstrap (spINAR-Bootstrap), the parametric (Poisson) INAR-Bootstrap (pINAR-Bootstrap), the Circular Block Bootstrap (CBB), the AR-Bootstrap (AR-Bootstrap), and the Markov Bootstrap (MB). We provide expressions for centering $\overline{X}^*$, $\widehat{\gamma}^*(0)$, $\widehat{\gamma}^*(1)$ and $\widehat{p}_0^*$; any of the remaining estimators is a function in $\overline{X}$, $\widehat{\gamma}(0)$, $\widehat{\gamma}(1)$, $\widehat{p}_0$, so we applied the corresponding function to $\mathrm{cent}(\overline{X}^*),\ldots$ to obtain an appropriate centering scheme.

spINAR

The semi-parametric estimation approach leads to the estimated (thinning) INAR coefficients $\widehat{\alpha}_{1;\mathrm{sp}}, \ldots, \widehat{\alpha}_{p;\mathrm{sp}}$ and innovations' probabilities $\widehat{g}_{\mathrm{sp}}(k)$ for $k = 0, \ldots, \max\{X_1, \ldots, X_n\}$. Using the latter, we compute $\widehat{\mu}_{\epsilon;\mathrm{sp}} := \sum_{k=0}^{\max\{X_1,\ldots,X_n\}} k \cdot \widehat{g}_{\mathrm{sp}}(k)$ and $\widehat{\sigma}^2_{\epsilon;\mathrm{sp}} := \sum_{k=0}^{\max\{X_1,\ldots,X_n\}} k^2 \cdot \widehat{g}_{\mathrm{sp}}(k) - \widehat{\mu}^2_{\epsilon;\mathrm{sp}}$. The probability generating function equals $\widehat{\mathrm{pgf}}_{\epsilon;\mathrm{sp}}(z) := \sum_{k=0}^{\max\{X_1,\ldots,X_n\}} z^k \cdot \widehat{g}_{\mathrm{sp}}(k)$. Centering of...

- $\overline{X}^*$: $\widehat{\mu}_{\epsilon;\mathrm{sp}}/(1 - \widehat{\alpha}_{1;\mathrm{sp}} - \cdots - \widehat{\alpha}_{p;\mathrm{sp}})$.
- $\widehat{\gamma}(0)^*$: $\mathrm{cent}(\overline{X}^*) \cdot \frac{\widehat{\alpha}_{1;\mathrm{sp}} + \frac{\widehat{\sigma}^2_{\epsilon;\mathrm{sp}}}{\widehat{\mu}_{\epsilon;\mathrm{sp}}}}{1+\widehat{\alpha}_{1;\mathrm{sp}}}$ for $p = 1$, and for $p = 2$:

$$\mathrm{cent}\big(\overline{X}^*\big) \cdot \frac{1 - \widehat{\alpha}^2_{1;\mathrm{sp}} - \widehat{\alpha}^2_{2;\mathrm{sp}} + (\frac{\widehat{\sigma}^2_{\epsilon;\mathrm{sp}}}{\widehat{\mu}_{\epsilon;\mathrm{sp}}} - 1)(1 - \widehat{\alpha}_{1;\mathrm{sp}} - \widehat{\alpha}_{2;\mathrm{sp}})}{1 - \widehat{\alpha}^2_{1;\mathrm{sp}} - \widehat{\alpha}^2_{2;\mathrm{sp}} - 2\widehat{\alpha}^2_{1;\mathrm{sp}} \frac{\widehat{\alpha}_{2;\mathrm{sp}}}{1-\widehat{\alpha}_{2;\mathrm{sp}}}}.$$

- $\widehat{\gamma}(1)^*$: $\mathrm{cent}(\widehat{\gamma}(0)^*) \cdot (1 - \frac{1}{n})\widehat{\alpha}_{1;\mathrm{sp}}$ for $p = 1$, and $\mathrm{cent}(\widehat{\gamma}(0)^*) \cdot (1 - \frac{1}{n})\frac{\widehat{\alpha}_{1;\mathrm{sp}}}{1-\widehat{\alpha}_{2;\mathrm{sp}}}$ for $p = 2$.
- $\widehat{p}_0^*$: For $p = 1$, we utilize a result in Jazi, Jones and Lai [20] to compute $\mathrm{cent}(\widehat{p}_0^*) := \prod_{j=0}^{M} \widehat{\mathrm{pgf}}_{\epsilon;\mathrm{sp}}(1 - \widehat{\alpha}^j_{1;\mathrm{sp}}) =: a(M)$, where M is the smallest value such that $|a(M) - a(M-1)| \leq 10^{-6}$.

  For $p = 2$, we ran an additional bootstrap loop and chose $\mathrm{cent}(\widehat{p}_0^*)$ as the mean zero frequency.

Note that (for simplicity) the centering for the sample variance and sample autocovariance is based on computations for $\widetilde{\gamma}^*(0) = \frac{1}{n}\sum_{t=1}^{n}(X_t^* - E^*(X_t^*))^2$ and $\widetilde{\gamma}^*(1) = \frac{1}{n}\sum_{t=1}^{n}(X_{t+1}^* - E^*(X_{t+1}^*))(X_t^* - E^*(X_t^*))$ instead of $\widehat{\gamma}^*(0)$ and $\widehat{\gamma}^*(1)$, respectively.

pINAR

The centerings for the pINAR bootstrap make use of Yule–Walker estimators $\overline{X}$ for the marginal mean μ_X and $\widehat{\alpha}_{1;\mathrm{YW}}, \ldots, \widehat{\alpha}_{p;\mathrm{YW}}$ (computed from the sample autocorrelation function according to the Yule–Walker equations of the INAR(p) model). Centering of...

- $\overline{X}^*$: $\overline{X}$.
- $\widehat{\gamma}(0)^*$: Because of the Poisson assumption concerning the innovations (equidispersion!), we choose again $\overline{X}$ for $p = 1$, and for $p = 2$:

$$\mathrm{cent}\big(\overline{X}^*\big) \cdot \frac{1 - \widehat{\alpha}^2_{1;\mathrm{YW}} - \widehat{\alpha}^2_{2;\mathrm{YW}}}{1 - \widehat{\alpha}^2_{1;\mathrm{YW}} - \widehat{\alpha}^2_{2;\mathrm{YW}} - 2\widehat{\alpha}^2_{1;\mathrm{YW}} \frac{\widehat{\alpha}_{2;\mathrm{YW}}}{1-\widehat{\alpha}_{2;\mathrm{YW}}}}.$$

- $\widehat{\gamma}(1)^*$: $\mathrm{cent}(\widehat{\gamma}(0)^*) \cdot (1 - \frac{1}{n})\widehat{\alpha}_{1;\mathrm{YW}}$ for $p = 1$, and $\mathrm{cent}(\widehat{\gamma}(0)^*) \cdot (1 - \frac{1}{n})\frac{\widehat{\alpha}_{1;\mathrm{YW}}}{1-\widehat{\alpha}_{2;\mathrm{YW}}}$ for $p = 2$.
- $\widehat{p}_0^*$: Because of the Poisson assumption, we choose $\exp(-\overline{X})$ for $p = 1$, while we ran an additional bootstrap loop for $p = 2$.

CBB

Centering of…

- $\overline{X}^*$: $\overline{X}$.
- $\widehat{\gamma}(0)^*$: $\widehat{\gamma}(0)$.
- $\widehat{\gamma}(1)^*$: $(1-\frac{\lceil n/b\rceil}{n})\frac{1}{n}\sum_{t=1}^{n}(X_{t+1}-\overline{X})(X_t-\overline{X})$, where we set $X_{n+1}=X_1$, and where b is the block length.
- $\widehat{p}_0^*$: $\widehat{p}_0$.

AR

Let $\widetilde{\epsilon}_t$ denote the centered AR residuals, see Section 2, and let $\widehat{\sigma}^2_{\epsilon;\mathrm{AR}} := \frac{1}{n-p}\sum_{t=p+1}^{n}\widetilde{\epsilon}_t^2$. Since for a continuous AR model, the probability of observing a zero equals 0, we treated the frequency of the event "$X_t < 0.5$" as the "zero frequency". Centering of…

- $\overline{X}^*$: $\overline{X}$.
- $\widehat{\gamma}(0)^*$: $\frac{\widehat{\sigma}^2_{\epsilon;\mathrm{AR}}}{1-\widehat{\alpha}^2_{1;\mathrm{YW}}}$ for $p=1$, and $\frac{\widehat{\sigma}^2_{\epsilon;\mathrm{AR}}}{1-\widehat{\alpha}^2_{1;\mathrm{YW}}-\widehat{\alpha}^2_{2;\mathrm{YW}}-2\widehat{\alpha}^2_{1;\mathrm{YW}}\frac{\widehat{\alpha}_{2;\mathrm{YW}}}{1-\widehat{\alpha}_{2;\mathrm{YW}}}}$ for $p=2$.
- $\widehat{\gamma}(1)^*$: $\mathrm{cent}(\widehat{\gamma}(0)^*)\cdot(1-\frac{1}{n})\widehat{\alpha}_{1;\mathrm{YW}}$ for $p=1$, and $\mathrm{cent}(\widehat{\gamma}(0)^*)\cdot(1-\frac{1}{n})\frac{\widehat{\alpha}_{1;\mathrm{YW}}}{1-\widehat{\alpha}_{2;\mathrm{YW}}}$ for $p=2$.
- $\widehat{p}_0^*$: We ran an additional bootstrap loop and chose $\mathrm{cent}(\widehat{p}_0^*)$ as the mean "zero frequency".

MB

Centering of…

- $\overline{X}^*$: $\overline{X}$.
- $\widehat{\gamma}(0)^*$: $\widehat{\gamma}(0)$.
- $\widehat{\gamma}(1)^*$: $\mathrm{cent}(\widehat{\gamma}(0)^*)\cdot(1-\frac{1}{n})\widehat{\alpha}_{1;\mathrm{YW}}$ for $p=1$, and $\mathrm{cent}(\widehat{\gamma}(0)^*)\cdot(1-\frac{1}{n})\frac{\widehat{\alpha}_{1;\mathrm{YW}}}{1-\widehat{\alpha}_{2;\mathrm{YW}}}$ for $p=2$.
- $\widehat{p}_0^*$: $\widehat{p}_0$.

E.3. Asymptotic approximations for $p = 1$

If the considered estimators are applied to a Poisson INAR(1) process, these estimators are asymptotically unbiased and normally distributed, with the asymptotic variances (inflated by n) given by

- $\overline{X}$: $\mu_X\frac{1+\alpha}{1-\alpha}$ (Weiß and Schweer [54]);
- $\widehat{\gamma}(0)$: $2\mu_X^2\frac{1+\alpha^2}{1-\alpha^2}+\mu_X\frac{1+\alpha}{1-\alpha}$ (Weiß and Schweer [54]);
- $\widehat{I}_{\mathrm{disp}}$: $2\frac{1+\alpha^2}{1-\alpha^2}$ (Schweer and Weiß [47]);
- $\widehat{\gamma}(1)$: $\mu_X^2\frac{1+4\alpha^2-\alpha^4}{1-\alpha^2}+\mu_X\frac{\alpha(1+\alpha)}{1-\alpha}$ (Weiß and Schweer [54]);
- $\widehat{\rho}(1)$: $1-\alpha^2+\frac{\alpha(1-\alpha)}{\mu_X}$ (Weiß and Schweer [54]);
- $\overline{X}(1-\widehat{\rho}(1))$: $\mu_\epsilon+\frac{1+\alpha}{1-\alpha}\mu_\epsilon^2$ (Weiß and Schweer [54]);
- $\widehat{p}_0$: $e^{-\mu_X}(1-e^{-\mu_X})+2e^{-2\mu_X}\sum_{j=1}^{\infty}\frac{\mu_X^j}{j!}\frac{\alpha^j}{1-\alpha^j}$ (Weiß, Puig and Homburg [53]);

Table 8. Semi-parametric INAR(p) estimates for time series of iceberg counts, see Appendix E.4

p	$\widehat{\alpha}_{1;\mathrm{sp}}$	$\widehat{\alpha}_{2;\mathrm{sp}}$	$\widehat{g}_{\mathrm{sp}}(0)$	$\widehat{g}_{\mathrm{sp}}(1)$	$\widehat{g}_{\mathrm{sp}}(2)$	$\widehat{g}_{\mathrm{sp}}(3)$	$\widehat{g}_{\mathrm{sp}}(4)$	$\widehat{g}_{\mathrm{sp}}(5)$	$\widehat{g}_{\mathrm{sp}}(6)$	$\widehat{g}_{\mathrm{sp}}(7)$	$\widehat{\mu}_{\epsilon;\mathrm{sp}}$	$\widehat{\sigma}^2_{\epsilon;\mathrm{sp}}$
1	0.528		0.698	0.250	0.043	0.005	0.002	0.001	0.000	0.000	0.368	0.405
2	0.474	0.162	0.760	0.205	0.029	0.003	0.003	0.000	0.000	0.000	0.283	0.313

- $\widehat{I}_{\mathrm{z.i.}}$: $\frac{e^{\mu_X}-1}{\mu_X^2} - \frac{1}{\mu_X}\frac{1+\alpha}{1-\alpha} + \frac{2}{\mu_X^2}\sum_{j=1}^{\infty}\frac{\mu_X^j}{j!}\frac{\alpha^j}{1-\alpha^j}$ (Weiß, Puig and Homburg [53]);
- $\widehat{I}_{\mathrm{z.m.}}$: $e^{\mu_X} - 1 - \mu_X\frac{1+\alpha}{1-\alpha} + 2\sum_{j=1}^{\infty}\frac{\mu_X^j}{j!}\frac{\alpha^j}{1-\alpha^j}$ (Weiß, Puig and Homburg [53]).

E.4. Additional results for data example

The time series of iceberg counts from Section 4.3 is of length $n = 1632$ and consists of values between 0 and 7. The sample estimates for the considered statistics are

$$\overline{x} \approx 0.778, \qquad \widehat{\gamma}(0) \approx 0.819, \qquad \widehat{I}_{\mathrm{disp}} \approx 1.053, \qquad \widehat{\gamma}(1) \approx 0.415, \qquad \widehat{\rho}(1) \approx 0.507,$$
$$\overline{x}\big(1-\widehat{\rho}(1)\big) \approx 0.383, \qquad \widehat{p}_0 \approx 0.463, \qquad \widehat{I}_{\mathrm{z.i.}} \approx 8.655 \cdot 10^{-3}, \qquad \widehat{I}_{\mathrm{z.m.}} \approx 6.752 \cdot 10^{-3}.$$

If using the semi-parametric approach of Drost, van den Akker and Werker [13] to fit an INAR(1) or INAR(2) model, with the initialization approach described in Appendix E.1, one obtains the estimates summarized in Table 8.

These estimates were used for the respective semi-parametric INAR(p) bootstrap.

The parametric INAR(p) bootstrap (Poisson innovations), and also the AR(p) bootstrap, used the following YW estimates: $\widehat{\alpha}_{1;\mathrm{YW}} \approx 0.507$, $\widehat{\mu}_{\epsilon;\mathrm{YW}} \approx 0.383$ for $p = 1$, and $\widehat{\alpha}_{1;\mathrm{YW}} \approx 0.412$, $\widehat{\alpha}_{2;\mathrm{YW}} \approx 0.188$, $\widehat{\mu}_{\epsilon;\mathrm{YW}} \approx 0.311$ for $p = 2$. As a result, the parametrically-fitted Poisson INAR models have slightly different properties than the semi-parametric ones as tabulated before (see Table 9).

The circular block bootstrap used block length 33.

Any bootstrap procedure was done with 10 000 replications. The obtained 95% confidence intervals are shown in Table 10.

Table 9. Parametric Poisson INAR(p) estimates for time series of iceberg counts, see Appendix E.4

p	$\widehat{\alpha}_{1;\mathrm{p}}$	$\widehat{\alpha}_{2;\mathrm{p}}$	$\widehat{g}_{\mathrm{p}}(0)$	$\widehat{g}_{\mathrm{p}}(1)$	$\widehat{g}_{\mathrm{p}}(2)$	$\widehat{g}_{\mathrm{p}}(3)$	$\widehat{g}_{\mathrm{p}}(4)$	$\widehat{g}_{\mathrm{p}}(5)$	$\widehat{g}_{\mathrm{p}}(6)$	$\widehat{g}_{\mathrm{p}}(7)$	$\widehat{\mu}_{\epsilon;\mathrm{p}}$	$\widehat{\sigma}^2_{\epsilon;\mathrm{p}}$
1	0.507		0.682	0.261	0.050	0.006	0.001	0.000	0.000	0.000	0.383	0.383
2	0.412	0.188	0.733	0.228	0.035	0.004	0.000	0.000	0.000	0.000	0.311	0.311

Table 10. 95% confidence intervals generated with different approaches (first lower bound, then upper bound) for time series of iceberg counts, see Appendix E.4

	p	spINAR		pINAR		CBB		AR		asymp	
$\bar{x}$	1	0.699	0.855	0.703	0.849	0.678	0.877	0.700	0.851	0.703	0.852
	2	0.670	0.878	0.683	0.871			0.682	0.868		
$\widehat{\gamma}(0)$	1	0.687	0.939	0.716	0.916	0.657	0.968	0.709	0.921	0.717	0.921
	2	0.625	0.990	0.661	0.961			0.706	0.922		
$\widehat{I}_{\text{disp}}$	1	0.933	1.162	0.962	1.139	0.913	1.186	0.915	1.179	0.964	1.142
	2	0.906	1.185	0.930	1.166			0.899	1.194		
$\widehat{\gamma}(1)$	1	0.314	0.505	0.333	0.491	0.291	0.524	0.337	0.489	0.334	0.496
	2	0.257	0.551	0.286	0.527			0.329	0.496		
$\widehat{\rho}(1)$	1	0.459	0.559	0.459	0.560	0.438	0.585	0.467	0.551	0.457	0.557
	2	0.451	0.571	0.448	0.577			0.460	0.560		
$\bar{x}(1-\widehat{\rho}(1))$	1	0.335	0.426	0.337	0.424	0.308	0.450	0.331	0.430	0.339	0.428
	2	0.329	0.432	0.325	0.433			0.318	0.439		
$\widehat{p}_0$	1	0.423	0.502	0.425	0.501	0.410	0.514	0.426	0.499	0.424	0.501
	2	0.416	0.510	0.419	0.506			0.418	0.508		
$\widehat{I}_{\text{z.i.}}$	1	−0.043	0.063	−0.040	0.059	−0.054	0.076	−0.043	0.060	−0.041	0.058
	2	−0.044	0.067	−0.042	0.064			−0.049	0.069		
$\widehat{I}_{\text{z.m.}}$	1	−0.035	0.049	−0.032	0.045	−0.044	0.059	−0.029	0.042	−0.032	0.045
	2	−0.043	0.055	−0.038	0.050			−0.034	0.048		

Acknowledgements

The authors thank the two referees for very useful comments on an earlier draft of this article. The iceberg order data of Section 4.3 were kindly made available to the second author by the Deutsche Börse. Prof. Dr. Joachim Grammig, University of Tübingen, is to be thanked for processing of it to make it amenable to data analysis. We are also very grateful to Prof. Dr. Robert Jung, University of Hohenheim, for his kind support to get access to the data. Large parts of this research was conducted while Carsten Jentsch held a position at the University of Mannheim, where he was financially supported by the German Research Foundation DFG via the Collaborative Research Center SFB 884 (Project B6) and the Baden-Württemberg-Stiftung via the Eliteprogram for Postdocs.

Supplementary Material

Supplement to "Bootstrapping INAR models" (DOI: 10.3150/18-BEJ1057SUPP; .pdf). We provide the full simulation results for Section 4.

References

[1] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **8** 261–275. MR0903755

[2] Alzaid, A.A. and Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *J. Appl. Probab.* **27** 314–324. MR1052303

[3] Basawa, I.V., Green, T.A., McCormick, W.P. and Taylor, R.L. (1990). Asymptotic bootstrap validity for finite Markov chains. *Comm. Statist. Theory Methods* **19** 1493–1510. MR1073045

[4] Bisaglia, L.B. and Gerolimetto, M. (2016). Estimation of INAR(p) models using bootstrap. Working Paper Series 06/2016, Univ. Padova.

[5] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR1093459

[6] Bu, R., McCabe, B. and Hadri, K. (2008). Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *J. Time Ser. Anal.* **29** 973–994. MR2464949

[7] Bühlmann, P. (1995). Bootstraps for time series. Technical Report 431, Dept. Statistics, Univ. California, Berkeley.

[8] Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148. MR1466304

[9] Cardinal, M., Roy, R. and Lambert, J. (1999). On the application of integer-valued time series models for the analysis of disease incidence. *Stat. Med.* **18** 2025–2039.

[10] De Schepper, A. and Heijnen, B. (1995). General restrictions on tail probabilities. *J. Comput. Appl. Math.* **64** 177–188. MR1364980

[11] Doukhan, P., Fokianos, K. and Li, X. (2012). On weak dependence conditions: The case of discrete valued processes. *Statist. Probab. Lett.* **82** 1941–1948. MR2970296

[12] Doukhan, P., Fokianos, K. and Li, X. (2013). Corrigendum to "On weak dependence conditions: The case of discrete valued processes" [Statist. Probab. Lett. 82 (2012) 1941–1948] [MR2970296]. *Statist. Probab. Lett.* **83** 674–675. MR3007006

[13] Drost, F.C., van den Akker, R. and Werker, B.J.M. (2009). Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 467–485. MR2649605

[14] Du, J.G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *J. Time Ser. Anal.* **12** 129–142. MR1108796

[15] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681

[16] Fink, T. and Kreiß, J.-P. (2013). Bootstrap for random coefficient autoregressive models. *J. Time Series Anal.* **34** 646–667. MR3127212

[17] Freeland, R.K. and McCabe, B. (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statist. Probab. Lett.* **73** 147–153. MR2159250

[18] Freeland, R.K. and McCabe, B.P.M. (2004). Analysis of low count time series data by Poisson autoregression. *J. Time Ser. Anal.* **25** 701–722. MR2089191

[19] Ibragimov, I.A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7** 349–382.

[20] Jazi, M.A., Jones, G. and Lai, C.-D. (2012). First-order integer valued AR processes with zero inflated Poisson innovations. *J. Time Series Anal.* **33** 954–963. MR2991911

[21] Jentsch, C. and Leucht, A. (2016). Bootstrapping sample quantiles of discrete data. *Ann. Inst. Statist. Math.* **68** 491–539. MR3489740

[22] Jentsch, C. and Politis, D.N. (2013). Valid resampling of higher-order statistics using the linear process bootstrap and autoregressive sieve bootstrap. *Comm. Statist. Theory Methods* **42** 1277–1293. MR3031281

[23] Jentsch, C. and Weiß, C.H. (2019). Supplement to "Bootstrapping INAR models." DOI:10.3150/18-BEJ1057SUPP.

[24] Johnson, N.L., Kemp, A.W. and Kotz, S. (2005). *Univariate Discrete Distributions*, 3rd ed. Hoboken, NJ: John Wiley & Sons. MR2163227

[25] Jung, R.C., McCabe, B.P.M. and Tremayne, A.R. (2016). Model validation and diagnostics. In *Handbook of Discrete-Valued Time Series*. *Chapman & Hall/CRC Handb*. *Mod*. *Stat*. *Methods* 189–218. Boca Raton, FL: CRC Press. MR3699406

[26] Jung, R.C. and Tremayne, A.R. (2006). Coherent forecasting in integer time series models. *Int*. *J*. *Forecast*. **22** 223–238.

[27] Jung, R.C. and Tremayne, A.R. (2011). Convolution-closed models for count time series with applications. *J*. *Time Series Anal*. **32** 268–280. MR2808254

[28] Kim, H.-Y. and Park, Y. (2006). Bootstrap confidence intervals for the INAR(p) process. *Korean Commun*. *Stat*. **13** 343–358.

[29] Kim, H.-Y. and Park, Y. (2008). A non-stationary integer-valued autoregressive model. *Statist*. *Papers* **49** 485–502. MR2399216

[30] Kreiß, J.-P. (1988). Asymptotical inference for a class of stochastic processes. Habilitationsschrift, Univ. Hamburg.

[31] Kreiß, J.-P. (1992). Bootstrap procedures for AR(∞)-processes. In *Bootstrapping and Related Techniques* (*Trier*, 1990). *Lecture Notes in Econom*. *and Math*. *Systems* **376** 107–113. Berlin: Springer. MR1238505

[32] Kreiß, J.-P. (1997). Asymptotical properties of residual bootstrap for autoregression. Preprint, TU Braunschweig.

[33] Kreiß, J.-P. and Paparoditis, E. (2011). Bootstrap methods for dependent data: A review. *J*. *Korean Statist*. *Soc*. **40** 357–378. MR2906623

[34] Kreiß, J.-P., Paparoditis, E. and Politis, D.N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann*. *Statist*. **39** 2103–2130. MR2893863

[35] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann*. *Statist*. **17** 1217–1241. MR1015147

[36] Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *J*. *Time Ser*. *Anal*. **19** 439–455. MR1652193

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

[38] Meintanis, S.G. and Karlis, D. (2014). Validation tests for the innovation distribution in INAR time series models. *Comput*. *Statist*. **29** 1221–1241. MR3266056

[39] Meyer, M., Jentsch, C. and Kreiß, J.-P. (2015). Baxter's inequality and sieve bootstrap for random fields. Working Paper, Univ. Mannheim.

[40] Meyer, M., Jentsch, C. and Kreiß, J.-P. (2017). Baxter's inequality and sieve bootstrap for random fields. *Bernoulli* **23** 2988–3020. MR3654797

[41] Meyer, M. and Kreiß, J.-P. (2015). On the vector autoregressive sieve bootstrap. *J*. *Time Series Anal*. **36** 377–397. MR3343007

[42] Park, Y. and Kim, H.-Y. (2012). Diagnostic checks for integer-valued autoregressive models using expected residuals. *Statist*. *Papers* **53** 951–970. MR2992927

[43] Patton, A., Politis, D.N. and White, H. (2009). Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White [MR2041534]. *Econometric Rev*. **28** 372–375. MR2532278

[44] Pavlopoulos, H. and Karlis, D. (2008). INAR(1) modeling of overdispersed count series with an environmental application. *Environmetrics* **19** 369–393. MR2440038

[45] Politis, D.N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev.* **23** 53–70. MR2041534

[46] Schweer, S. (2016). A goodness-of-fit test for integer-valued autoregressive processes. *J. Time Series Anal*. **37** 77–98. MR3439533

[47] Schweer, S. and Weiß, C.H. (2014). Compound Poisson INAR(1) processes: Stochastic properties and testing for overdispersion. *Comput*. *Statist*. *Data Anal*. **77** 267–284. MR3210062

[48] Silva, I. and Silva, M.E. (2006). Asymptotic distribution of the Yule–Walker estimator for INAR(p) processes. *Statist*. *Probab. Lett*. **76** 1655–1663. MR2248854

[49] Steutel, F.W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Ann*. *Probab.* **7** 893–899. MR0542141

[50] Tsay, R.S. (1992). Model checking via parametric bootstraps in time series analysis. *Appl*. *Stat*. **41** 1–15.

[51] Weiß, C.H. (2012). Process capability analysis for serially dependent processes of Poisson counts. *J*. *Stat*. *Comput*. *Simul*. **82** 383–404. MR2897647

[52] Weiß, C.H. (2018). *An Introduction to Discrete-Valued Time Series*. Chichester: Wiley.

[53] Weiß, C.H., Puig, P. and Homburg, A. (2016). Testing for zero inflation in INAR(1) models. *Statist*. *Papers*. To appear. DOI:10.1007/s00362-016-0851-y.

[54] Weiß, C.H. and Schweer, S. (2016). Bias corrections for moment estimators in Poisson INAR(1) and INARCH(1) processes. *Statist*. *Probab. Lett*. **112** 124–130. MR3475496

[55] Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch*. *Verw*. *Gebiete* **52** 45–57. MR0568258

Received September 2017 and revised March 2018