

Beran, Jan; Droullier, Frieder

Article — Published Version

On random coefficient INAR processes with long memory

AStA Advances in Statistical Analysis

Suggested Citation: Beran, Jan; Droullier, Frieder (2025) : On random coefficient INAR processes with long memory, AStA Advances in Statistical Analysis, ISSN 1863-818X, Springer Berlin Heidelberg, Berlin/Heidelberg, Vol. 109, Iss. 2, pp. 281-311, <https://doi.org/10.1007/s10182-025-00523-8>

This Version is available at:

<https://hdl.handle.net/10419/323287>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



On random coefficient INAR processes with long memory

Jan Beran¹ · Frieder Droullier¹

Received: 31 July 2024 / Accepted: 3 January 2025 / Published online: 24 January 2025
© The Author(s) 2025

Abstract

We consider random coefficient INAR(1) processes with a strongly dependent latent random coefficient process. It is shown that, in spite of its conditional Markovian structure, the unconditional process exhibits long-range dependence. Short-term prediction and estimation of parameters involved in the prediction are considered. Asymptotic rates of convergence are derived.

Keywords Integer valued time series · INAR process · Long memory · Long-range dependence · Random coefficient process · Gaussian subordination

1 Introduction

We consider integer valued time series defined by an INAR(1) equation with a randomly generated autoregressive parameter. INAR(1) processes based on the binomial thinning operator (Steutel and van Harn 1979) were originally introduced in McKenzie (1985) and Al-Osh and Alzaid (1987). By definition, INAR(1) processes are Markovian. In this paper, the autoregressive parameter is assumed to vary as a function of time. Random coefficient INAR(1) processes with iid coefficients ϕ_t are considered in Leonenko et al. (2007) and Zheng et al. (2007). Also see Cui and Wang (2019) for random coefficient INAR(1) models based on a negative binomial thinning operator. The iid assumption preserves the Markov property. In contrast, in the present paper, the latent coefficient process ϕ_t ($t \in \mathbb{Z}$) is assumed to exhibit long-range dependence. Under suitable technical assumptions, this implies long-range dependence in the observed process X_t . In particular, the Markov property no longer holds. The second question addressed in this paper is prediction of X_{t+1} given X_t and estimation of $\phi_t(r) = E(\phi_t | X_{t-1} = r)$ ($r \geq 0$) and $\lambda_Z = E(X_t | X_{t-1} = 0)$. Consistency and asymptotic rates of convergence are derived.

There is an extensive literature on INAR(p) and related processes. For an overview see, e.g., Weiß (2018). A very partial list of relevant references includes for instance Du

✉ Jan Beran
jan.beran@uni-konstanz.de

¹ Department of Mathematics and Statistics, University of Konstanz, Konstanz, Germany

and Li (1991), Gauthier and Latour (1994), Freeland and McCabe (2005), Silva et al. (2009), Pedeli and Karlis (2011), Schweer and Weiß (2014), Bourguignon et al. (2016), and Jentsch and Weiß (2019). Long memory generated by aggregating INAR(1) processes is studied in Leonenko et al. (2007) and Barczy et al. (2017). Beran and Droullier (2023) consider strongly dependent zero-inflated INAR(1) processes. Also, Beran (2002) and Jia et al. (2023) study the generation of integer valued processes with long memory by Gaussian subordination. Quoreschi (2014) discusses the possibility of defining INARFIMA models with long-range dependence. However, the results in Quoreschi (2014) are not correct, because linear filters with non-summable coefficients cannot be carried over directly to thinning filters. Applying such filters to non-negative sequences leads to infinite values. The processes in Quoreschi (2014) are therefore not defined. This problem was recognized by Segnon and Stapper (2019). They use truncated fractional filters, thereby obtaining higher order Markov processes that can be used as a heuristic approximation of long memory. Strictly speaking however, these processes do not exhibit long-range dependence. For a general review on count series, including processes with long-range dependence, see Davis et al. (2016, 2021). General references to long memory processes can be found for instance in Beran (1994), Giraitis et al. (2012), Beran et al. (2013) and Pipiras and Taqqu (2017). Empirically, the presence of long memory in count data has been observed in many applications (Hurvich et al. 2009; Hainaut and Boucher 2014; Braccini 2015). For instance, in finance, long memory in processes characterizing durations between transactions often lead to long memory in count processes (see, e.g., Daley et al. 2000; Hurvich et al. 2009).

The paper is organized as follows. Basic definitions are given in section 2. The existence of a stationary solution is derived in section 3. The asymptotic behavior of the autocovariance function is investigated in section 4. Prediction and parameter estimation are considered in section 5. Simulations and data examples in section 6 illustrate the results. Proofs are given in the Appendix.

2 Basic definitions

Let $X \geq 0$ be an integer valued random variable and $\phi \in (0, 1)$. The thinning operator " \circ " is defined by

$$\phi \circ X = \sum_{i=1}^X \xi_i \quad (1)$$

where ξ_i are iid Bernoulli random variables with $P(\xi_i = 1) = \phi$ (Steutel and van Harn 1979). Thus, conditionally on X , $Y = \phi \circ X$ is a binomial random variable with parameters X and ϕ . An INAR(1) process (McKenzie 1985; Al-Osh and Alzaid 1987) is defined by

$$X_t = \phi \circ X_{t-1} + Z_t \quad (t \in \mathbb{Z}), \quad (2)$$

where $Z_t \geq 0$ are iid integer valued random variables with $E(Z_t) = \lambda_Z$.

In this paper, we consider a random coefficient INAR(1) process defined by

$$X_t = \phi_t \circ X_{t-1} + Z_t \quad (t \in \mathbb{Z}), \quad (3)$$

where ϕ_t ($t \in \mathbb{Z}$) is a second order stationary process with $\phi_t \in (0, 1)$, $E(\phi_t) = \phi$ and autocovariance function $\gamma_\phi(k) = \text{cov}(\phi_t, \phi_{t+k})$ ($k \in \mathbb{Z}$). More specifically, we consider the case where

$$\gamma_\phi(k) \sim c_\phi |k|^{2d-1} \quad (k \rightarrow \infty) \quad (4)$$

for some $0 < c_\phi < \infty$ and $d \in (0, \frac{1}{2})$. Here, " \sim " means that the ratio of the left and right hand side tends to one. Equation (4) implies

$$\sum_{k=-\infty}^{\infty} \gamma_\phi(k) = \infty. \quad (5)$$

Stationary processes with property (5) are said to have strong dependence, long memory, or long-range dependence.

The following assumptions will be used:

- (A1) X_t is defined by (3).
- (A2)

$$\phi_t \in (0, b)$$

for some $b = 1 - \epsilon$, $0 < \epsilon < 1$, and $E(\phi_t) = \phi$.

- (A3) The process $(\phi_t)_{t \in \mathbb{Z}}$ is strictly stationary with autocovariance function γ_ϕ satisfying (4).
- (A4) The processes $(\phi_t)_{t \in \mathbb{Z}}$ and $(Z_t)_{t \in \mathbb{Z}}$ are independent of each other. Also, conditionally on (ϕ_t, X_{t-1}) , the binomial random variable $\phi_t \circ X_{t-1}$ is independent of ϕ_s ($s \leq t-1$), X_s ($s \leq t-2$), and Z_s ($s \leq t$).
- (A5) For all $s_1, s_2, t_1, t_2 \in \mathbb{Z}$, $s_1 \leq s_2$, $t_1 \leq t_2$,

$$\text{cov}\left(\prod_{j=s_1}^{s_2} \phi_j, \prod_{j=t_1}^{t_2} \phi_j\right) \geq 0.$$

- (A6) Let

$$p(r) = P(X_0 = r),$$

$$p(k; r, s) = P(X_0 = r, X_k = s),$$

and

$$\begin{aligned} \gamma_1(k; r, s) &= \text{cov}(1\{X_0 = r\} 1\{X_k = s\}) \\ &= p(k; r, s) - p(r)p(s). \end{aligned}$$

Then,

$$\lim_{k \rightarrow \infty} k^{1-2d} \gamma_1(k; r, s) \sim c_{\gamma,1}(r, s) k^{2d-1} \quad (6)$$

for some $0 < c_{\gamma,1}(r, s) < \infty$.

3 Stationarity

Using the convention that, for $j_1 > j_2$, $\prod_{j=j_1}^{j_2} \phi_j = 1$, we define

$$X_{t,n} = \sum_{i=0}^n \left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i}. \quad (7)$$

Under (A1) to (A4), a stationary solution of (3) exists and can be obtained as an almost sure limit of $X_{t,n}$:

Theorem 1 *Suppose that (A1) to (A4) hold. Then, for each $t \in \mathbb{Z}$, there exists a random variable X_t with $E(X_t^2) < \infty$ such that $X_{t,n}$ converges to X_t almost surely. We then write*

$$X_t = \sum_{i=0}^{\infty} \left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i}. \quad (8)$$

Moreover, X_t ($t \in \mathbb{Z}$) is strictly stationary, and the unique solution of (3).

Remark 1 Note that

$$E(X_t) = E \left\{ E \left[\sum_{i=0}^{\infty} \left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i} \mid \phi_s, s \leq t \right] \right\} \quad (9)$$

$$= \lambda_Z \sum_{i=0}^{\infty} E \left(\prod_{j=t-i+1}^t \phi_j \right). \quad (10)$$

Remark 2 A strongly dependent coefficient process satisfying (A2) and (A3) can be obtained for instance by Gaussian subordination. Let ζ_t be a stationary Gaussian process with $E(\zeta_t) = 0$, $\text{var}(\zeta_t) = 1$ and autocovariance function $\gamma_{\zeta}(k) \sim c_{\gamma,\zeta} |k|^{2d-1}$ where $d \in (0, \frac{1}{2})$. We define

$$\phi_t = \phi + G(\zeta_t) \quad (11)$$

where $G : \mathbb{R} \rightarrow (a, b) \subseteq (-\phi, 1 - \phi)$ is such that $E[G(\zeta)] = 0$, $E[G^2(\zeta)] < \infty$. If $E[\zeta G(\zeta)] \neq 0$, then the function G is said to have Hermite rank one, and the autocovariance function of ϕ_t is of the form

$$\gamma_{\phi}(k) \sim c_{\gamma,\phi} |k|^{2d-1} \quad (12)$$

for some $0 < c_{\gamma,Y} < \infty$. Thus, the long memory property of ζ_t carries over to ϕ_t . The process $\phi_t = \phi + G(\zeta_t)$ is said to be generated by Gaussian subordination. For references and a more general discussion of Gaussian subordination see, e.g., Taqqu (1975), Major (1981), Dobrushin and Major (1979), Giraitis et al. (2012), Beran et al. (2013), Pipiras and Taqqu (2017), and references therein.

Example 1 Let ζ_t be defined as in Remark 2, $\phi \in \left(\frac{1}{4}, \frac{3}{4}\right)$, $0 < c < \frac{1}{4}$, $a > 0$, and

$$\phi_t = \phi + c \tanh(a\zeta_t)$$

Then with probability one, $\phi_t \in (0, 1)$ and (12) holds.

4 Long-range dependence

The autocovariance function of random coefficient INAR(1) processes with iid coefficients is of the form $\gamma_X(k) = \gamma_X(0)\phi^k$ ($k \geq 0$) (Leonenko et al. 2007; Zheng et al. 2007). Since $0 < E(\phi_t) = \phi < 1$, $\gamma_X(k)$ is summable. Here we consider assumptions (A1) to (A4), and in particular γ_ϕ of the form (4). Sufficient conditions will be given that imply a non-summable autocovariance function γ_X .

We will use the convention that an empty product is equal to one, i.e., $\prod_{j=r}^s \phi_j = 1$ ($s < r$). First note that (8) implies the following representation of $\gamma_X(k)$:

Theorem 2 Suppose that (A1) to (A4) hold. Then

$$\gamma_X(k) = \gamma_C(k) + c_2(k) \quad (13)$$

where

$$\begin{aligned} \gamma_C(k) &= \text{cov}(C_t, C_{t+k}), \\ C_t &= \lambda_Z \sum_{i=0}^{\infty} \prod_{j=t-i+1}^t \phi_j \end{aligned} \quad (14)$$

and

$$c_2(k) = \sigma_Z^2 \sum_{i=0}^{\infty} E \left(\prod_{j=t-i+1}^t \phi_j \prod_{j=t-i+1}^{t+k} \phi_j \right).$$

Due to stationarity, c_2 and C_t can also be written as

$$c_2(k) = \sigma_Z^2 \sum_{i=0}^{\infty} E \left(\prod_{j=1}^i \phi_j \prod_{j=1}^{i+k} \phi_j \right) = \sigma_Z^2 \sum_{i=0}^{\infty} E \left(\prod_{j=1}^i \phi_j^2 \prod_{j=i+1}^{i+k} \phi_j \right),$$

and

$$\begin{aligned}
 \gamma_C(k) &= \text{cov} \left(\lambda_Z \sum_{i=0}^{\infty} \prod_{j=t-i+1}^t \phi_j, \lambda_Z \sum_{i=0}^{\infty} \prod_{j=t+k-i+1}^{t+k} \phi_j \right) \\
 &= \lambda_Z^2 \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} \text{cov} \left(\prod_{j=1-i_1}^0 \phi_j, \prod_{j=1-i_2+k}^k \phi_j \right).
 \end{aligned} \tag{15}$$

Assumption (A2) leads to an exponential decay of $c_2(k)$:

Lemma 1 Suppose that $\phi_t \in (0, 1 - \epsilon)$ for some $\epsilon > 0$. Then there are constants $0 < K < \infty$, $\beta \in (0, 1)$ and an integer $k_0 \geq 1$ such that

$$|c_2(k)| \leq K\beta^k \quad (k \geq k_0).$$

A lower bound for the autocovariance function can be given as follows:

Theorem 3 Under assumptions (A1) to (A5), there is a constant $0 < C_\gamma < \infty$ and a $k_0 \in \mathbb{N}$ such that

$$\gamma_X(k) \geq C_\gamma k^{2d-1} \quad (k \geq k_0). \tag{16}$$

As a special case we may consider random variables with positive quadrant dependence (Lehmann 1966), or association as defined in Esary et al. (1967). We recall the following definitions.

Definition 1 (Lehmann 1966) Two real valued random variables X and Y are called positively quadrant dependent, if for all $x, y \in \mathbb{R}$,

$$P(X \leq x, Y \leq y) \geq P(X \leq x)P(Y \leq y).$$

Definition 2 (Esary et al. 1967) Denote by $\mathbf{Y} = (Y_1, \dots, Y_m)$ a vector of real valued random variables Y_j . Then Y_j ($j = 1, \dots, m$) are called associated, if for all nondecreasing functions f and g ,

$$\text{cov}(f(\mathbf{Y}), g(\mathbf{Y})) \geq 0,$$

provided that the covariance is well defined.

Corollary 1 Let $\eta_t = \log \phi_t$. Suppose that for any $m \geq 1$, the random variables η_0, \dots, η_m are associated. Then the lower bound (16) holds.

Remark 3 Suppose that $\gamma_\zeta(k) \geq 0$ for all k . Then the assumptions of Corollary 1 can be achieved for instance by Gaussian subordination (see Remark 2) where G is a suitable monotonically increasing function.

Remark 4 Theorem 3 provides sufficient conditions such that the process X_t has a non-summable autocovariance function. In addition to a lower bound for γ_X it would be desirable to obtain an asymptotic expression of the form

$$\gamma_X(k) \sim c_{\gamma,X} k^{2d-1} \quad (k \rightarrow \infty)$$

($0 < c_{\gamma,X} < \infty$) together with an explicit formula for the constant $c_{\gamma,X}$. Deriving an explicit formula for $c_{\gamma,X}$ appears to be difficult in general. For instance, under Gaussian subordination $\phi_t = \phi + G(\zeta_t)$, one may use the Hermite expansion of $G(\zeta)$ and diagram formulas for products of Hermite polynomials (Major 1981; Giraitis and Surgailis 1986; Malyshev and Minlos 1991; Surgailis 2003; also see Beran et al. 2013 and references therein). Combinatorial questions in this context are subject of current research.

Remark 5 Note that, conditionally on ϕ_t and X_{t-1} , the observation X_t is independent of the past X_s ($s < t$). Nevertheless, unconditionally X_t exhibits long-range dependence. The reason is that long-range dependence in the coefficient process ϕ_t leads to strongly dependent fluctuations in the distribution of X_t . For instance, $E[X_t | \phi_t, X_{t-1}] = \phi_t X_{t-1} + \lambda$. As a result, unconditionally, long memory in ϕ_t carries over to the process X_t . Note in particular that, unlike cross sectional aggregation of linear AR(1) models (see, e.g., Granger 1980), this effect has nothing to do with the range the random coefficients ϕ_t .

5 One step ahead prediction

We consider optimal prediction of X_{t+1} given X_t . The mean squared prediction error is minimized by

$$\begin{aligned} \hat{X}_{t+1} &= E(X_{t+1} | X_t) = E(\phi_{t+1} \circ X_t + Z_t | X_t) \\ &= E(\phi_{t+1} | X_t) \cdot X_t + \lambda_Z. \end{aligned}$$

Thus, for $r = 0, 1, 2, \dots$, let

$$\phi(r) = E(\phi_{t+1} | X_t = r).$$

Then

$$\hat{X}_{t+1} = E(X_{t+1} | X_t = r) = \phi(r) \cdot r + \lambda_Z$$

and

$$X_{t+1} - \hat{X}_{t+1} = (Y_{t+1} - \phi(r) \cdot r) + (Z_{t+1} - \lambda_Z),$$

where conditionally on $X_t = r$, Y_{t+1} is independent of Z_{t+1} and $Y_{t+1} \sim \text{Bin}(r, \phi(r))$. Note in particular that

$$E(X_{t+1} | X_t = 0) = \lambda_Z.$$

The conditional expected quadratic prediction error is equal to

$$\begin{aligned}\text{MSPE}(r) &= E\left[(X_{t+1} - \hat{X}_{t+1})^2 \mid X_t = r\right] \\ &= \phi(r)(1 - \phi(r)) \cdot r + \sigma_Z^2.\end{aligned}$$

In practice, λ_Z and $\phi(r)$ ($r \geq 1$) are unknown. Simple consistent estimators of λ_Z and $\phi(r)$ can be given as follows. Let

$$\hat{p}_n(r) = \frac{1}{n} \sum_{t=2}^n 1\{X_{t-1} = r\},$$

and define

$$\hat{\lambda}_Z = \frac{n^{-1} \sum_{t=2}^n 1\{X_{t-1} = 0\} X_t}{\hat{p}_n(0)}. \quad (17)$$

Moreover note that, for $r \geq 1$,

$$X_{t+1} - \lambda_Z = Y_{t+1} + (Z_{t+1} - \lambda_Z),$$

so that

$$\begin{aligned}E\left(\frac{X_{t+1} - \lambda_Z}{r} \mid X_t = r\right) &= E\left(\frac{Y_{t+1}}{r} \mid X_t = r\right) \\ &= E(\phi_{t+1} \mid X_t = r) = \phi(r).\end{aligned} \quad (18)$$

If $p(r) = P(X_t = r) > 0$, then (18) motivates to define the estimator

$$\hat{\phi}(r) = \frac{1}{nr} \frac{S_n(r)}{\hat{p}_n(r)} \quad (19)$$

where

$$S_n(r) = \sum_{t=2}^n \xi_t(r),$$

and

$$\xi_t(r) = 1\{X_{t-1} = r\} (X_t - \hat{\lambda}_Z).$$

In cases where $\hat{p}_n(r) = 0$ we set $\hat{\phi}(r)$ equal to zero. Consistency and asymptotic rates of convergence of $\hat{\lambda}_Z$ and $\hat{\phi}(r)$ are given in the following Theorem.

Theorem 4 Suppose that (A1) to (A6) hold, and $p(r) > 0$. Then, as $n \rightarrow \infty$,

$$\begin{aligned}\hat{\lambda}_Z &\xrightarrow{p} \lambda_Z, \quad \hat{\phi}(r) \xrightarrow{p} \phi(r) \\ E(\hat{\lambda}_Z) &= \lambda_Z + o\left(n^{-\frac{1}{2}}\right), \quad E(\hat{\phi}(r)) = \phi(r) + o\left(n^{-\frac{1}{2}}\right),\end{aligned}$$

and

$$\lim_{n \rightarrow \infty} \text{cov} \left(\sqrt{n}(\hat{\lambda}_Z - \lambda_Z), \sqrt{n}(\hat{\phi}(r) - \phi(r)) \right) = 0.$$

Moreover, there are constants $0 < c_{\text{est},\lambda}, c_{\text{est},\phi}(r) < \infty$ such that

$$\begin{aligned} \text{var}(\hat{\lambda}_Z) &\sim c_{\text{est},\lambda} n^{-1}, \\ \text{var}(\hat{\phi}(r)) &\sim c_{\text{est},\phi}(r) n^{-1}. \end{aligned}$$

Remark 6 Theorem 4 implies that the mean square errors of $\hat{\lambda}_Z$ and $\hat{\phi}(r)$ are asymptotically of the form

$$\text{MSE}(\hat{\lambda}_Z) \sim \text{var}(\hat{\lambda}_Z) \sim c_{\text{est},\lambda} n^{-1}$$

and

$$\text{MSE}(\hat{\phi}(r)) \sim \text{var}(\hat{\phi}(r)) \sim c_{\text{est},\phi}(r) n^{-1}.$$

It is quite remarkable that, in spite of long-range dependence in the observed process, both estimation errors are of the order $O_p(n^{-1/2})$.

Remark 7 In exactly the same way, we may estimate

$$\phi^* = E(\phi_t | X_{t-1} \neq 0) \quad (20)$$

by

$$\hat{\phi}^* = \frac{n^{-1} \sum_{t=2}^n 1\{X_{t-1} \neq 0\} (X_t - \hat{\lambda}_Z) / X_{t-1}}{n^{-1} \sum_{t=2}^n 1\{X_{t-1} \neq 0\}}. \quad (21)$$

An analogous result as in Theorem 4 can be derived.

Remark 8 Estimation of the unconditional expected value $\phi = E(\phi_t)$ is more difficult. The reason is that for $X_{t-1} = 0$, $X_t = Z_t$ does not include any information about ϕ_t . Thus, even if the noise process Z_t were known completely, X_t would not provide any information about $\phi(0) = E(\phi_t | X_{t-1} = 0)$. In the case of iid coefficients ϕ_t (Leonenko et al. 2007; Zheng et al. 2007), this problem does not occur, because ϕ_t is independent of X_{t-1} .

Remark 9 Note that $\hat{\phi}(r)$ is a semiparametric estimator, in the sense that (3) is assumed, but no other specifications are needed. Long memory in the latent process ϕ_t suggests that forecast accuracy may be improved by taking into account a larger number of past observations. Thus, given observations X_1, \dots, X_t , one may consider predictions

$$\hat{X}_{t+1}^{(m)} = E(X_{t+1} | X_{t-s}, 0 \leq s \leq m)$$

where $0 \leq m \leq t-1$. Analogous formulas involving λ_Z and conditional expected values of the coefficient process can be obtained. However, when m is large, then the probability that the vector $(X_{t-s}, 0 \leq s \leq m)$ assumes a specific value may become very small. Therefore, the application of a semiparametric estimator analogous to (19) becomes problematic. In particular, for $m = t-1$, such an estimator is not consistent. Instead one would need to express $\hat{X}_{t+1}^{(m)}$ in terms of the model parameters ϕ , λ_Z , $c_{\gamma, \phi}$ and d , and apply a fully parametric estimator. This is a formidable task worth pursuing in future research.

Remark 10 With respect to prediction of count time series, some authors recommend to use so-called coherent forecast approaches, such as the conditional median, see Homburg et al. (2019). A possible approach that combines this idea with estimates of λ and $\phi(r)$ given here is discussed in the following section.

6 Simulations and data examples

6.1 Simulations

A small simulation study illustrates Theorem 4. We consider model (3) with iid Poisson distributed variables Z_t , and

$$\phi_t = \phi + \frac{1}{10} \tanh(10\zeta_t),$$

where ζ_t is a FARIMA(0, d , 0) process with $\text{var}(\zeta_t) = 1$ (Granger and Joyeux 1980; Hosking 1981). The following parameter settings are considered for each $d = 0.1, 0.2, 0.3$ and 0.4 : a) $\phi = 0.5$, $E(Z_t) = \lambda_Z = \sigma_Z^2 = 2$, and b) $\phi = 0.5$, $\lambda_Z = 2.5$. The range of $\mu = E(X_t)$ is $5.0 \leq \mu \leq 5.1$ in case a), and $6.3 \leq \mu \leq 6.4$ in case b). Figure 1a shows typical sample paths of X_t for $\lambda = 2$. A bar chart with simulated probabilities $\hat{p}(r)$ based on the fourth series ($d = 0.4$) is shown in Fig. 1b.

For each parameter constellation, and sample sizes $n = 800, 1000, 1200, 1400, 1600, 1800$ and 2000 , $N = 10'000$ sample paths were simulated. For each sample path, $\hat{\lambda}$, $\hat{\phi}^*$, $\hat{\phi}(1)$ and $\hat{\phi}(5)$ were calculated, as defined in (17), (21) and (19), respectively. Table 1 shows the simulated means and variances of the estimates for case a). For illustration, the logarithm of the simulated variances of $\hat{\lambda}$, $\hat{\phi}(1)$ and $\hat{\phi}^*$ are plotted against $\log n$ in Fig. 2a, b and c for $\lambda = 2$. For $\log(\text{var}(\hat{\lambda}))$ vs. $\log n$, the slopes of the fitted least squares lines are very close to the theoretical value of minus one: -1.18 ($d = 0.1$), -1.25 ($d = 0.2$), -1.30 ($d = 0.3$), -1.47 ($d = 0.4$). Similar results are obtained for $\log(\text{var}(\hat{\phi}(1)))$ ($-0.82, -0.80, -0.79, -0.75$), $\log(\text{var}(\hat{\phi}(5)))$ ($-1.25, -1.13, -1.12, -1.12$) and $\log(\text{var}(\hat{\phi}^*))$ ($-1.24, -1.16, -1.18, -1.17$). These results illustrate the expected \sqrt{n} -rate of convergence of the estimators. Comparable results are obtained for case b) (Table 2).

The formula for $\hat{\lambda}$ given in (17) has the advantage of being simple. However, $\hat{\lambda}$ may be highly volatile when $p(0)$ is small. Comparing Tables 1 and 2, the simulations illustrate that $\text{var}(\hat{\lambda})$ increases considerably in spite of the rather minor

Table 1 Simulated means (first number) and variances (second number) of $\hat{\lambda}$, $\hat{\phi}^*$, $\hat{\phi}(1)$ and $\hat{\phi}(5)$, for Case a) ($\lambda = 2$)

n	$\hat{\lambda}_Z$	$\hat{\phi}^*$	$\hat{\phi}(1)$	$\hat{\phi}(5)$
$d = 0.1$				
800	2.079, 0.435	0.581, 0.028	0.634, 0.282	0.584, 0.018
1000	2.063, 0.321	0.584, 0.021	0.616, 0.240	0.587, 0.013
1200	2.052, 0.249	0.587, 0.017	0.604, 0.207	0.589, 0.010
1400	2.044, 0.206	0.589, 0.014	0.596, 0.181	0.591, 0.009
1600	2.039, 0.173	0.590, 0.012	0.590, 0.162	0.592, 0.007
1800	2.034, 0.150	0.591, 0.010	0.590, 0.147	0.593, 0.006
2000	2.030, 0.134	0.592, 0.009	0.589, 0.134	0.594, 0.006
$d = 0.2$				
800	2.067, 0.414	0.584, 0.026	0.638, 0.280	0.586, 0.017
1000	2.046, 0.308	0.589, 0.020	0.626, 0.237	0.591, 0.013
1200	2.035, 0.239	0.591, 0.016	0.614, 0.207	0.593, 0.010
1400	2.031, 0.195	0.592, 0.013	0.605, 0.181	0.594, 0.009
1600	2.028, 0.170	0.593, 0.012	0.600, 0.162	0.594, 0.007
1800	2.024, 0.149	0.594, 0.010	0.597, 0.147	0.595, 0.007
2000	2.019, 0.132	0.595, 0.009	0.595, 0.136	0.596, 0.006
$d = 0.3$				
800	2.080, 0.443	0.582, 0.027	0.624, 0.281	0.583, 0.018
1000	2.056, 0.316	0.587, 0.020	0.608, 0.234	0.588, 0.014
1200	2.042, 0.245	0.590, 0.016	0.601, 0.205	0.591, 0.011
1400	2.033, 0.198	0.592, 0.014	0.594, 0.180	0.593, 0.009
1600	2.029, 0.172	0.593, 0.012	0.590, 0.163	0.594, 0.008
1800	2.023, 0.150	0.594, 0.010	0.589, 0.147	0.595, 0.007
2000	2.021, 0.134	0.595, 0.009	0.586, 0.136	0.595, 0.006
$d = 0.4$				
800	2.118, 0.607	0.575, 0.032	0.624, 0.297	0.576, 0.024
1000	2.074, 0.419	0.584, 0.024	0.613, 0.255	0.584, 0.018
1200	2.053, 0.323	0.588, 0.019	0.604, 0.224	0.588, 0.014
1400	2.039, 0.253	0.591, 0.016	0.597, 0.199	0.591, 0.012
1600	2.032, 0.212	0.593, 0.014	0.592, 0.178	0.592, 0.010
1800	2.027, 0.178	0.594, 0.012	0.587, 0.162	0.594, 0.009
2000	2.024, 0.159	0.595, 0.011	0.585, 0.150	0.594, 0.008

increase in λ from 2 to 2.5. The reason is that this increase lowers $p(0)$ by a factor of about 1/10. This leads to much larger simulated variances of $\hat{\lambda}$. An improvement of (17) may be achieved by including non-zero values of X_{t-1} as well, for instance by using appropriate combinations of conditional log-likelihood functions and/or equations for moments. Typically, formulas of this type are however no longer explicit. A detailed investigation of optimal estimation procedures and their asymptotic and finite sample properties is an interesting question that should be addressed by future research.

Table 2 Simulated means (first number) and variances (second number) of $\hat{\lambda}$, $\hat{\phi}^*$, $\hat{\phi}(1)$ and $\hat{\phi}(5)$, for Case b) ($\lambda = 2.5$)

n	$\hat{\lambda}_Z$	$\hat{\phi}^*$	$\hat{\phi}(1)$	$\hat{\phi}(5)$
$d = 0.1$				
800	2.967, 1.859	0.510, 0.070	0.672, 0.642	0.508, 0.072
1000	2.816, 1.566	0.540, 0.059	0.702, 0.593	0.538, 0.060
1200	2.713, 1.288	0.560, 0.049	0.712, 0.545	0.558, 0.050
1400	2.665, 1.083	0.568, 0.042	0.701, 0.495	0.567, 0.042
1600	2.618, 0.904	0.577, 0.035	0.698, 0.459	0.576, 0.035
1800	2.588, 0.792	0.583, 0.031	0.695, 0.426	0.582, 0.031
2000	2.572, 0.698	0.586, 0.027	0.689, 0.401	0.585, 0.028
$d = 0.2$				
800	2.956, 1.835	0.513, 0.069	0.684, 0.667	0.508, 0.071
1000	2.809, 1.529	0.542, 0.058	0.708, 0.612	0.537, 0.059
1200	2.727, 1.281	0.557, 0.048	0.707, 0.550	0.553, 0.049
1400	2.670, 1.061	0.568, 0.040	0.700, 0.499	0.564, 0.041
1600	2.624, 0.867	0.576, 0.034	0.689, 0.449	0.573, 0.034
1800	2.600, 0.752	0.581, 0.029	0.682, 0.417	0.577, 0.030
2000	2.579, 0.644	0.585, 0.025	0.673, 0.386	0.581, 0.026
$d = 0.3$				
800	2.939, 1.819	0.518, 0.067	0.681, 0.643	0.509, 0.070
1000	2.808, 1.525	0.543, 0.056	0.699, 0.587	0.535, 0.058
1200	2.713, 1.259	0.561, 0.046	0.707, 0.538	0.553, 0.049
1400	2.658, 1.072	0.571, 0.040	0.704, 0.499	0.564, 0.041
1600	2.622, 0.918	0.578, 0.034	0.696, 0.458	0.571, 0.036
1800	2.597, 0.793	0.582, 0.030	0.688, 0.426	0.576, 0.031
2000	2.579, 0.694	0.586, 0.026	0.679, 0.392	0.579, 0.027
$d = 0.4$				
800	3.030, 1.853	0.507, 0.061	0.630, 0.614	0.488, 0.068
1000	2.902, 1.586	0.530, 0.052	0.645, 0.561	0.513, 0.058
1200	2.803, 1.365	0.547, 0.045	0.657, 0.523	0.533, 0.051
1400	2.737, 1.178	0.559, 0.039	0.659, 0.484	0.545, 0.044
1600	2.686, 1.015	0.568, 0.034	0.660, 0.453	0.555, 0.039
1800	2.651, 0.890	0.574, 0.030	0.655, 0.418	0.562, 0.034
2000	2.621, 0.778	0.579, 0.027	0.652, 0.394	0.568, 0.031

A further question of interest is the quality of one-step ahead forecasts given X_{t-1} . By definition, conditionally on X_{t-1} and ϕ_t , the random variable X_t is independent of the past. Therefore, instead of point forecasts it is more meaningful to predict conditional quantiles of X_t , given X_{t-1} . For illustration, we consider the conditional median, say M , of X_t given $X_{t-1} = r$. The prediction results for $r = 5$ are summarized in Tables 3 and 4. For all parameter settings considered here, M is equal to 5. The results in Tables 3 and 4 are based on $N = 10'000$ simulations. Given N simulated series of length n , estimated conditional medians $\hat{M}_{i,n}$ ($i = 1, 2, \dots, N$) were calculated. Here, $\hat{M}_{i,n}$ is defined as the median of

Table 3 Simulated distribution of estimated predicted medians for Case a) ($\lambda = 2$): Let M denote the conditional median of X_t given $X_{t-1} = 5$. The table displays simulated distributions of estimates of M . For each sample size n , $N = 10^4$ series were simulated and estimates of M , denoted by $\hat{M}_{i,n}$ ($i = 1, 2, \dots, N$), were calculated. The table shows relative frequencies $f_{n,M}(m) = N^{-1} \sum_{i=1}^N 1\{\hat{M}_{i,n} = m\}$ rounded to three digits after the decimal point

n	$f_{n,M}(4)$	$f_{n,M}(5)$	$f_{n,M}(6)$
$d = 0.1$			
800	0.006	0.994	0
1200	0	1	0
1600	0	1	0
2000	0	1	0
$d = 0.2$			
800	0.008	0.991	0.001
1200	0.001	0.999	0
1600	0	1	0
2000	0	1	0
$d = 0.3$			
800	0.016	0.982	0.002
1200	0.006	0.994	0
1600	0.002	0.998	0
2000	0.001	0.999	0
$d = 0.4$			
800	0.061	0.932	0.006
1200	0.042	0.956	0.002
1600	0.030	0.968	0.001
2000	0.022	0.977	0.001

Table 4 Simulated distribution of estimated predicted medians for Case b) ($\lambda = 2.5$): Let M denote the conditional median of X_t given $X_{t-1} = 5$. The table displays simulated distributions of estimates of M . For each sample size n , $N = 10^4$ series were simulated and estimates of M , denoted by $\hat{M}_{i,n}$ ($i = 1, 2, \dots, N$), were calculated. The table shows relative frequencies $f_{n,M}(m) = N^{-1} \sum_{i=1}^N 1\{\hat{M}_{i,n} = m\}$ rounded to three digits after the decimal point

n	$f_{n,M}(5)$	$f_{n,M}(6)$	$f_{n,M}(7)$	$f_{n,M}(8)$
$d = 0.1$				
800	0.744	0.246	0.003	0.001
1200	0.780	0.212	0.002	0.001
1600	0.808	0.185	0.001	0
2000	0.830	0.163	0.001	0
$d = 0.2$				
800	0.750	0.243	0.002	0
1200	0.787	0.209	0.002	0
1600	0.815	0.183	0.001	0
2000	0.843	0.157	0	0
$d = 0.3$				
800	0.753	0.234	0.002	0.001
1200	0.788	0.207	0.002	0.001
1600	0.813	0.184	0.002	0
2000	0.830	0.168	0.001	0
$d = 0.4$				
800	0.725	0.266	0.003	0
1200	0.744	0.249	0.002	0
1600	0.755	0.241	0.001	0
2000	0.762	0.236	0	0

Table 5 Microsoft—estimated conditional medians given $X_{t-1} = r$ ($r = 1, 2, 3$). The results are based on a series of length $N = 2000$. The estimates (first number in each column) are based on the first n observations X_1, \dots, X_n and compared to the corresponding conditional sample medians of the future observations X_{n+1}, \dots, X_N (second number in each column)

n	$X_{t-1}=1$	2	3
600	1, 1	1, 1	2, 2
800	1, 1	1, 2	2, 1
1000	1, 1	2, 1	2, 2
1200	1, 1	1, 1	2, 1.5
1400	1, 1	1, 1	2, 1

Table 6 Netflix—estimated conditional medians given $X_{t-1} = r$ ($r = 1, 2, \dots, 6$). The results are based on a series of length $N = 2000$. The estimates (first number in each column) are based on the first n observations X_1, \dots, X_n and compared to the corresponding conditional sample medians of the future observations X_{n+1}, \dots, X_N (second number in each column)

n	$X_{t-1}=1$	2	3	4	5	6
600	1, 1	2, 2	3, 3	3, 3	5, 3	4, 4
800	1, 1	2, 2	3, 3	3, 3	4, 3.5	4, 3.5
1000	1, 1	2, 2	3, 3	3, 3	4, 3	4, 3.5
1200	1, 1	2, 2	3, 3	3, 3	4, 3.5	4, 3.5
1400	1, 1	2, 2	3, 3	3, 3	4, 4	4, 3.5

$\hat{X}_t = \hat{Y} + \hat{Z}$ where \hat{Y} and \hat{Z} are mutually independent, \hat{Z} is Poisson distributed with intensity $\hat{\lambda}$ and \hat{Y} is a binomial random variable \hat{Y} with $r = 5$ Bernoulli trials and success probability $\hat{\phi}(5)$. Tables 3 and 4 show the simulated distributions of $\hat{M}_{i,n}$ for all sample sizes and parameter settings. Relative frequencies are denoted by $f_{n,M}(m) = N^{-1} \sum_{i=1}^N 1\{\hat{M}_{i,n} = m\}$. The results show an increased concentration of $\hat{M}_{i,n}$ around the correct value of $M = 5$, as n increases. The probability of obtaining a correct estimate of the conditional median is generally close to one, in particular when long memory is weak ($d=0.1, 0.2$). A comparison of the two cases indicates however that increasing λ makes prediction of M more difficult. This is not unexpected in view of the increased variances of $\hat{\lambda}$ and $\hat{\phi}(5)$ (see Tables 1 and 2; Figs. 3 and 4).

6.2 Data examples

Transaction counts in high-frequency trading tend to exhibit long memory (see, e.g., Hurvich et al. 2009). Though most high-frequency data are not in the public domain, the presence of long memory can be detected by examining suitable aggregates that are freely available. This is due to the fact that temporal aggregation preserves long-range dependence (see, e.g., Chapter 2.2.1 in Beran

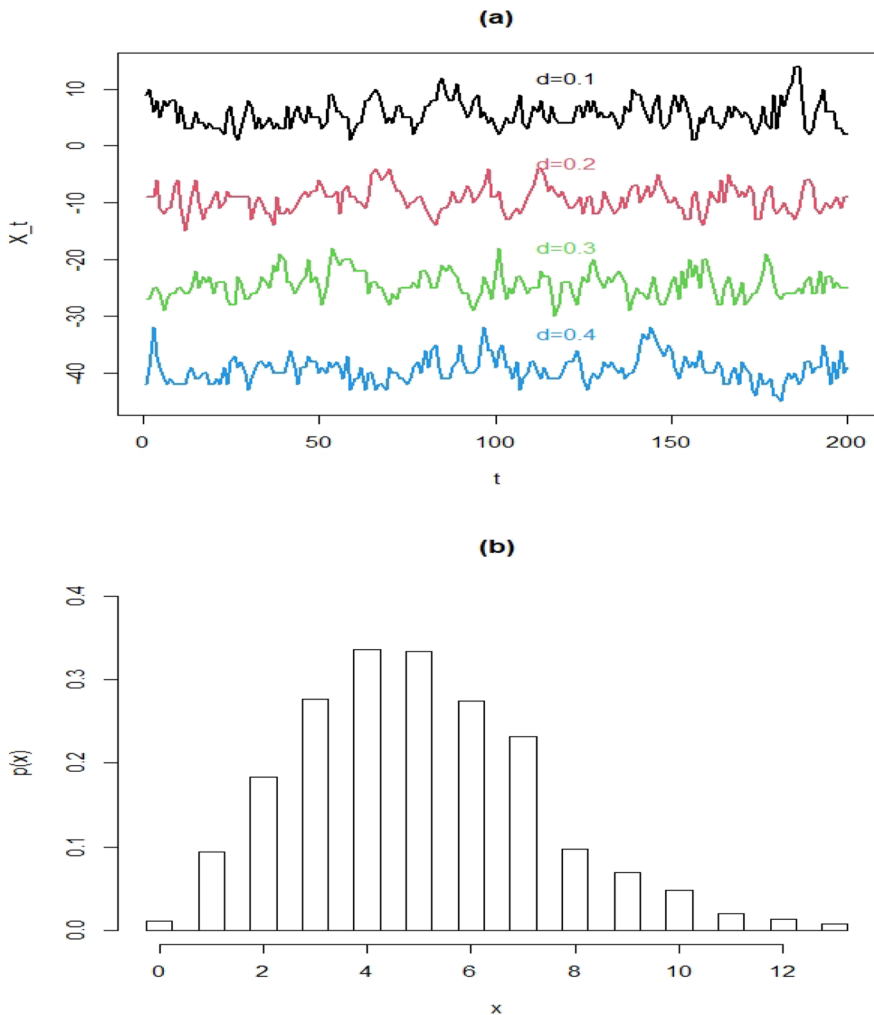


Fig. 1 Simulated sample paths of an RCINAR(1) process with $d = 0.1, 0.2, 0.3$ and 0.4 , respectively, **a**. For better visibility, the series are shifted vertically. **b** Shows a bar chart of simulated relative frequencies $\hat{p}(x)$ ($x = 0, 1, 2, \dots$) for the fourth series ($d = 0.4$)

et al. 2013, and references therein). Figures 3a and 5a show daily trading volumes for Microsoft and Netflix, for a time span of about eight years (July 29, 2016 to July 11, 2024). The data were downloaded from the Nasdaq public domain platform (<https://www.nasdaq.com/>). Histograms of the series are given in Figs. 3d and 5d, respectively. The slow decay of the sample autocorrelations (Figs. 3b and 5b) and the negative slopes of log-log-periodograms (Figs. 3c and 5c) are clear indications for the presence of long memory. Note also that there is one extreme outlier in the Netflix series. The outlier is due to a selloff

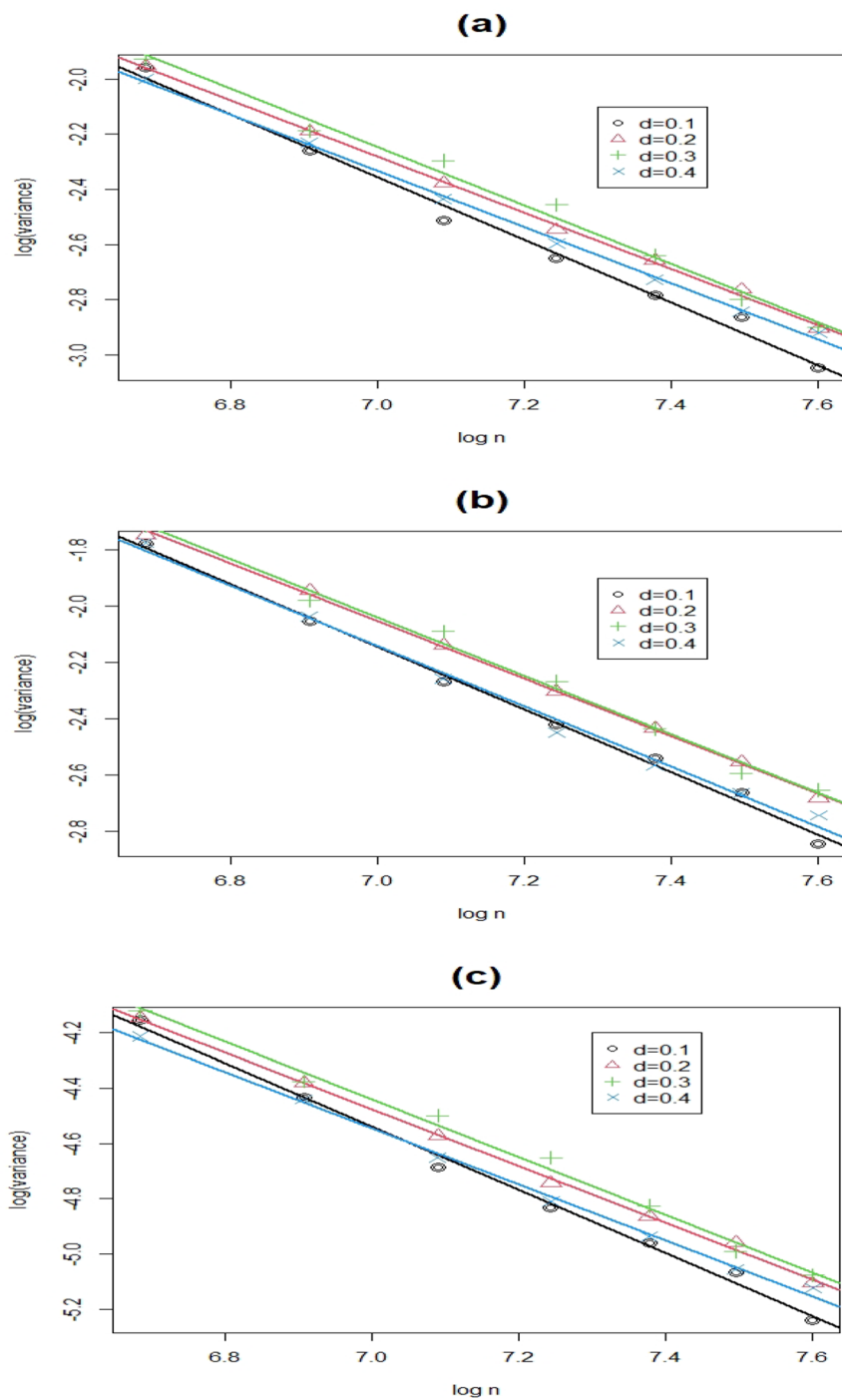


Fig. 2 Logarithm of simulated variances of $\hat{\lambda}_Z$ (a), $\hat{\phi}(1)$ (b) and $\hat{\phi}^*$ (c) plotted against log n

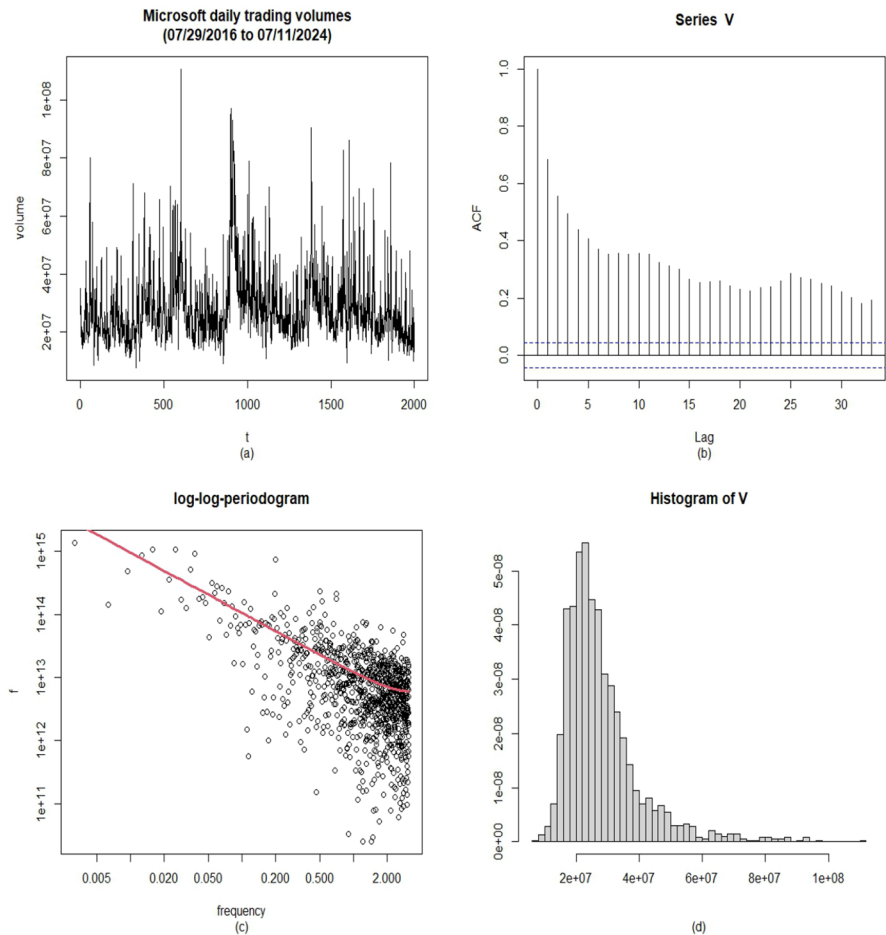


Fig. 3 Microsoft: daily trading volumes (a), sample autocorrelations (b), log-log-periodogram (c) and histogram (d)

on April 20, 2022, after the company's announcement of a loss of 200'000 subscribers in the first quarter of the year.

Next we illustrate prediction of conditional medians. Let $\Delta t > 0$ be measured in seconds, and denote by V_t the volume on day t . Nasdaq is open for six and a half hours a day. The average number of transactions in a time span of Δt seconds is therefore equal to $X_t^0 = \Delta t \cdot V_t / (6.5 \cdot 3600)$. We define X_t by rounding X_t^0 to the nearest integer. Since Microsoft tends to have much higher volumes than Netflix, we use different values of Δt (0.001 s for Microsoft, 0.005 s for Netflix). Figures 4a and 6a display X_t for the two companies. The correlograms, log-log-periodograms and histograms are given in Figs. 4b, c, d and 6b, c, d, respectively. Both series are of length $N = 2000$. Tables 5 and 6 display estimated conditional medians given $X_{t-1} = r$, based on the first n observations X_1, \dots, X_n , where $n = 600, 800, 1000, 1200$ and 1400. The estimated conditional

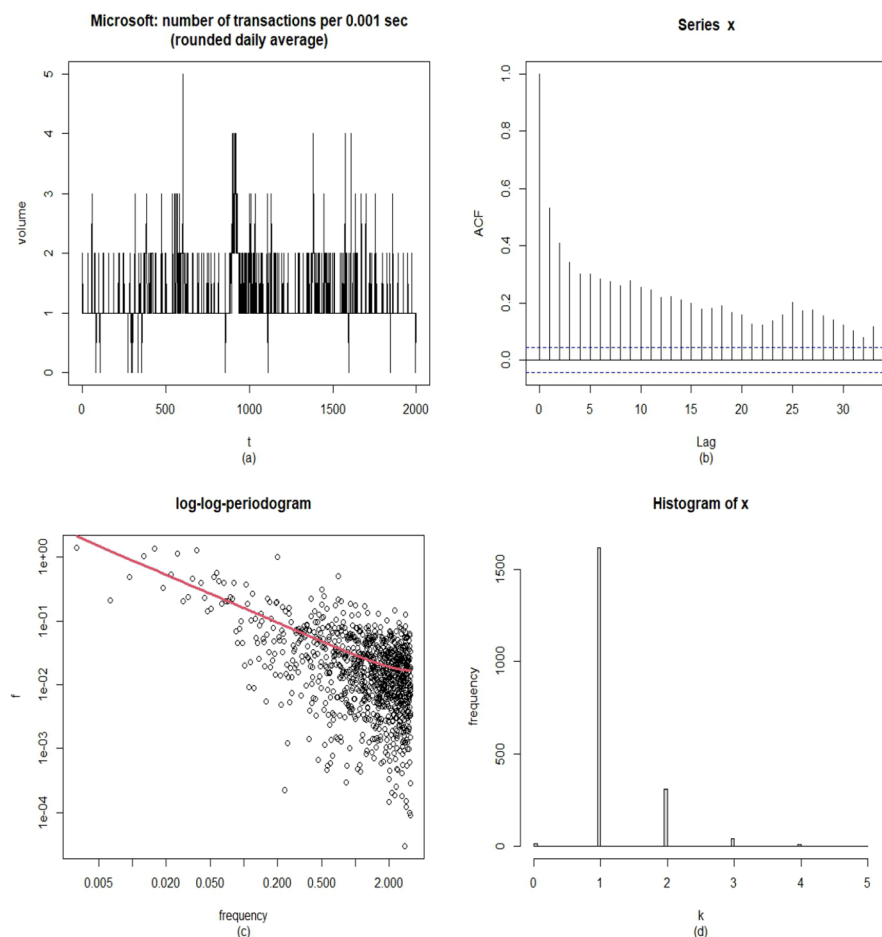


Fig. 4 Microsoft: daily average number of transactions per 0.001 s (a), sample autocorrelations (b), log-log-periodogram (c) and histogram (d)

medians are compared to the corresponding conditional sample medians of the future observations X_{n+1}, \dots, X_N . Taking into account that the sample medians are estimates as well, the results show a reasonably good agreement between the two quantities.

7 Final remarks

In this paper we introduced random coefficient INAR(1) processes generated by a strongly dependent stationary parameter sequence ϕ_t . A lower bound for the autocovariance function of the INAR(1) process was derived, implying non-summable

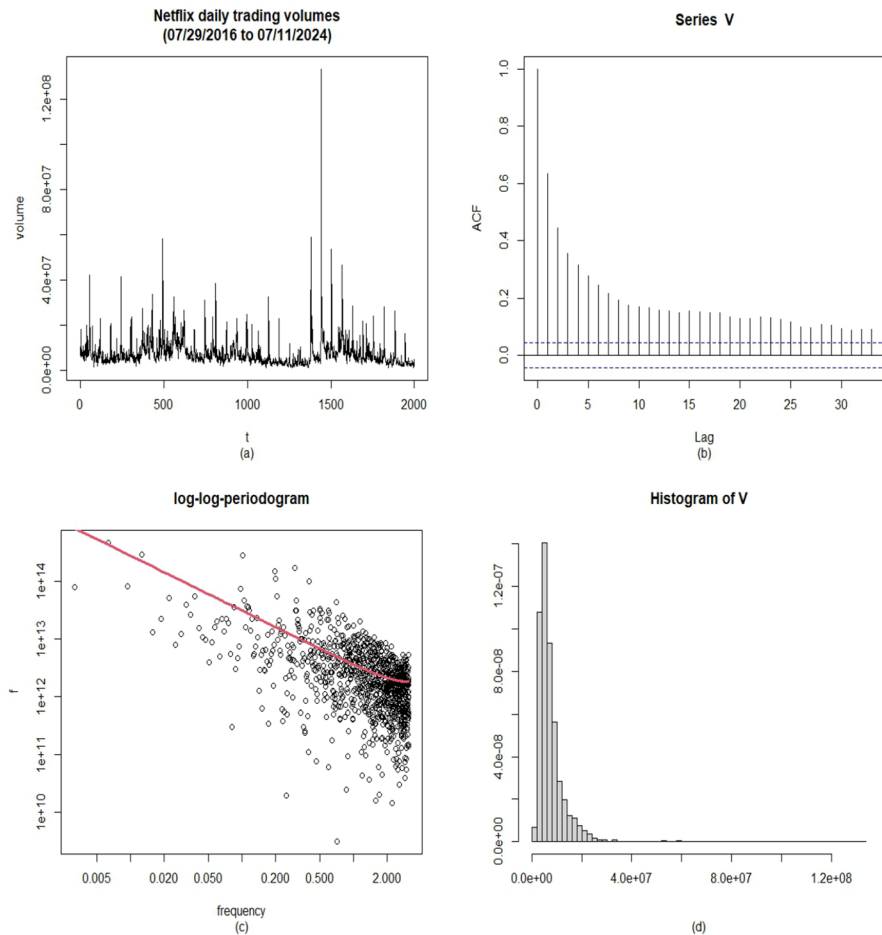


Fig. 5 Netflix: daily trading volumes (a), sample autocorrelations (b), log-log-periodogram (c) and histogram (d)

autocorrelations. One-step-ahead prediction given the immediate past was considered. Consistent estimators of parameters needed for prediction were proposed and their asymptotic rate of convergence was derived.

Many interesting open questions remain, including a detailed study of the asymptotic dependence structure and other distributional properties, improved forecasts, and an extension to strongly dependent random coefficient INAR(p) processes with $p \geq 2$. Also, the development of more efficient parameter estimators should be addressed. In particular, $\hat{\lambda}$ as defined in (17) may be too volatile in cases where $X_t = 0$ is a rare event.

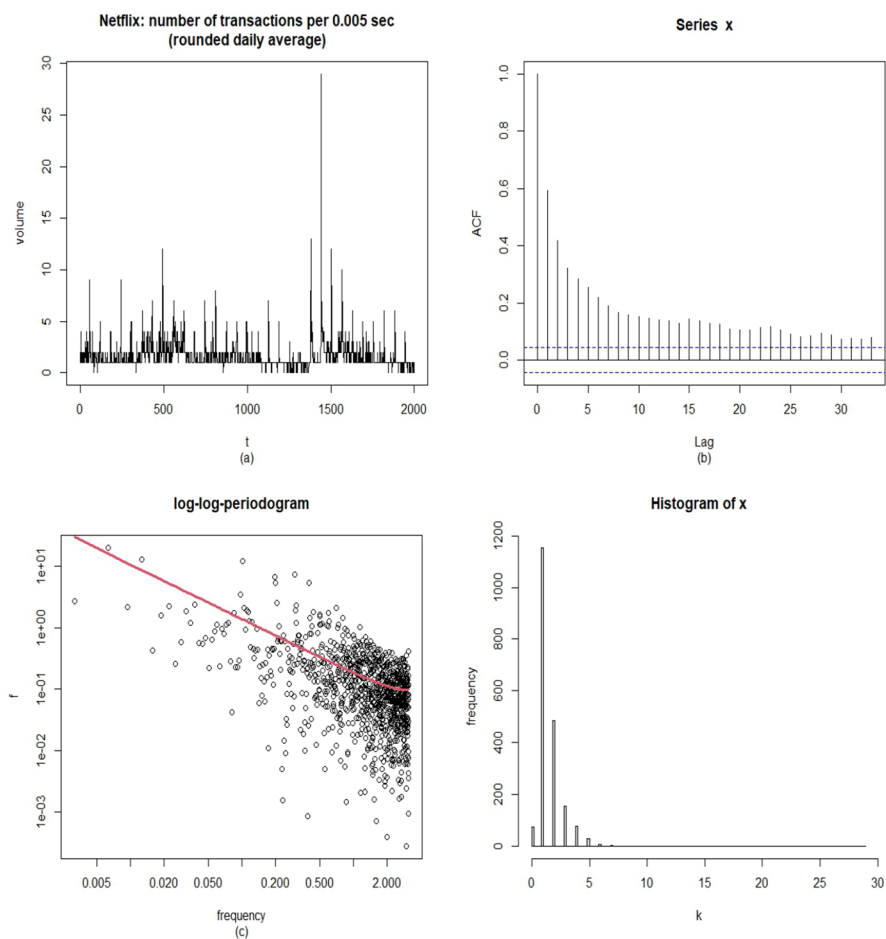


Fig. 6 Netflix: daily average number of transactions per 0.005 s **(a)**, sample autocorrelations **(b)**, log-log-periodogram **(c)** and histogram **(d)**

Appendix

Proofs

Proof (of Theorem 1) First, we prove the existence of X_t as an almost sure limit of $X_{t,n}$. Note that $\phi_t \in [0, b]$ with $b < 1$, implies

$$E\left(\prod_{j=t-i+1}^t \phi_j\right) = E\left(\prod_{j=1}^i \phi_j\right) \leq b^i.$$

Let

$$X_{t,n} = \sum_{i=0}^n \xi_{t,i}$$

where

$$\xi_{t,i} = \left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i}.$$

Note that, by definition, $\xi_{t,i} \geq 0$ so that

$$0 \leq X_{t,n} = \sum_{i=0}^n \xi_{t,i} = \sum_{i=0}^n |\xi_{t,i}| \leq \sum_{i=0}^{n+1} |\xi_{t,i}| = X_{t,n+1}. \quad (22)$$

To prove almost sure convergence of $X_{t,n}$ we can now apply the same arguments as in Proposition 13.3.1 in Brockwell and Davis (1991). Due to (22), convergence of $E(|X_{t,n}|) = E(X_{t,n})$ implies almost sure convergence of $X_{t,n}$. Now, $E(|X_{t,n}|)$ is monotonically nondecreasing, with an upper bound given by

$$\begin{aligned} E(|X_{t,n}|) &= E(X_{t,n}) \\ &= E \left[\sum_{i=0}^n \left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i} \right] \\ &= \sum_{i=0}^n E \left\{ E \left[\left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i} \mid \phi_s, s \leq t \right] \right\} \\ &= \lambda_Z \sum_{i=0}^n E \left(\prod_{j=1}^i \phi_j \right) \leq \lambda_Z \sum_{i=0}^n b^i \leq \frac{\lambda_Z}{1-b} < \infty. \end{aligned}$$

Therefore, $E(|X_{t,n}|)$ converges to a finite value, and hence $X_{t,n}$ converges almost surely to

$$X_t = \sum_{i=0}^{\infty} \left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i}. \quad (23)$$

Note also that $E(|X_t|) < \infty$. Strict stationarity of X_t follows from strict stationarity of ϕ_t .

To obtain weak stationarity, we need to prove $E(X_t^2) < \infty$. Setting $a_i = \prod_{j=t-i+1}^t \phi_j \circ Z_{t-i}$, the inequality

$$E(Z_i Z_j) \leq \lambda_Z^2 + \sigma_Z^2, \quad E(Z_i^2) = \lambda_Z^2 + \sigma_Z^2,$$

implies

$$\sum_{i_1=0}^{\infty} a_{i_1} \sum_{i_2=0}^{\infty} a_{i_2} = \sum_{i=0}^{\infty} \sum_{k=0}^i a_k a_{i-k} \leq (\lambda_Z^2 + \sigma_Z^2) \sum_{i=0}^{\infty} (i+1)b^i, \quad (24)$$

and hence

$$\begin{aligned} E[X_t^2] &\leq (\lambda_Z^2 + \sigma_Z^2) E \left[\sum_{i=0}^{\infty} \left(\prod_{j=t-i+1}^t \phi_j \right) \sum_{i=0}^{\infty} \left(\prod_{j=t-i+1}^t \phi_j \right) \right] \\ &\leq (\lambda_Z^2 + \sigma_Z^2) \sum_{i=0}^{\infty} (i+1)b^i = (\lambda_Z^2 + \sigma_Z^2) \frac{1}{(1-b)^2} < \infty. \end{aligned}$$

Next we show that X_t is a solution of (3). Recall the semigroup property and distributivity of binomial thinning operators (Scotto et al. 2015):

$$\alpha \circ (\beta \circ X) \stackrel{d}{=} \beta \circ (\alpha \circ X) \stackrel{d}{=} (\alpha \beta) \circ X, \quad (25)$$

$$\alpha \circ (Z + Y) \stackrel{d}{=} \alpha \circ Z + \alpha \circ Y \quad (26)$$

for any $\alpha, \beta \in [0, 1]$, and integer valued random variables $X, Y, Z \geq 0$. Now, let X_t be defined by (23). Then, using (25) and (26),

$$\begin{aligned} \phi_{t+1} \circ X_t + Z_{t+1} &= \phi_{t+1} \circ \left(\sum_{i=0}^{\infty} \left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i} \right) + Z_{t+1} \\ &= \left(\sum_{i=0}^{\infty} \left(\phi_{t+1} \prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i} \right) + Z_{t+1} \\ &= \left(\sum_{i=0}^{\infty} \left(\prod_{j=t-i+1}^{t+1} \phi_j \right) \circ Z_{t-i} \right) + Z_{t+1} \\ &= \left(\sum_{i=1}^{\infty} \left(\prod_{j=(t+1)-i+1}^{t+1} \phi_j \right) \circ Z_{(t+1)-i} \right) + Z_{t+1}. \end{aligned}$$

Now,

$$\left(\prod_{j=(t+1)+1}^{t+1} \phi_j \right) \circ Z_{t+1} = 1 \circ Z_{t+1} = Z_{t+1},$$

so that

$$\phi_{t+1} \circ X_t + Z_{t+1} = \sum_{i=0}^{\infty} \left(\prod_{j=(t+1)-i+1}^{t+1} \phi_j \right) \circ Z_{t+1-i} = X_{t+1}.$$

Uniqueness of X_t follows by analogous arguments. \square

Proof (of Theorem 2) Using the notation

$$C_t = \lambda_Z \sum_{i=0}^{\infty} \prod_{j=t-i+1}^t \phi_j$$

and the representation

$$X_t = \sum_{i=0}^{\infty} \left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i},$$

we have

$$\begin{aligned} E(X_t) &= \sum_{i=0}^{\infty} E \left[E \left(\left(\prod_{j=t-i+1}^t \phi_j \right) \circ Z_{t-i} \mid \phi_s, s \leq t \right) \right] \\ &= \sum_{i=0}^{\infty} E \left[\lambda_Z \prod_{j=t-i+1}^t \phi_j \right] = E(C_t). \end{aligned}$$

Since

$$\begin{aligned} &E \left[\left(\prod_{j=t-i_1+1}^t \phi_j \right) \circ Z_{t-i_1} \cdot \left(\prod_{j=t+k-i_2+1}^{t+k} \phi_j \right) \circ Z_{t+k-i_2} \mid \phi_s, Z_s, s \leq t+k \right] \\ &= \prod_{j=t-i_1+1}^t \phi_j Z_{t-i_1} \prod_{j=t+k-i_2+1}^{t+k} \phi_j Z_{t+k-i_2}, \end{aligned}$$

we have

$$\begin{aligned} \gamma_X(k) + E^2(C_0) &= \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} E \left(\prod_{j=t-i_1+1}^t \phi_j \prod_{j=t+k-i_2+1}^{t+k} \phi_j \cdot Z_{t-i_1} Z_{t+k-i_2} \right) \\ &= \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} E(Z_{t-i_1} Z_{t+k-i_2}) E \left(\prod_{j=t-i_1+1}^t \phi_j \prod_{j=t+k-i_2+1}^{t+k} \phi_j \right). \end{aligned}$$

Since

$$\begin{aligned} E(Z_{t-i_1} Z_{t+k-i_2}) &= \lambda_Z^2 (i_2 \neq i_1 + k), \\ E(Z_{t-i_1} Z_{t+k-i_2}) &= \sigma_Z^2 + \lambda_Z^2 (i_2 = i_1 + k), \end{aligned}$$

we obtain

$$\begin{aligned}
& \gamma_X(k) + E^2(C_0) \\
&= \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} E(Z_{t-i_1} Z_{t+k-i_2}) E\left(\prod_{j=t-i_1+1}^t \phi_j \prod_{j=t+k-i_2+1}^{t+k} \phi_j\right) \\
&= \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \lambda_Z^2 E\left(\prod_{j=t-i_1+1}^t \phi_j \prod_{j=t+k-i_2+1}^{t+k} \phi_j\right) \\
&\quad + \sum_{i=0}^{\infty} \sigma_Z^2 E\left(\prod_{j=t-i+1}^t \phi_j \prod_{j=t-i+1}^{t+k} \phi_j\right) \\
&= E(C_t C_{t+k}) + c_2(k).
\end{aligned}$$

Thus

$$\gamma_X(k) = \text{cov}(C_t, C_{t+k}) + c_2(k) = \gamma_C(k) + c_2(k).$$

□

Proof (of Lemma 1) Assumption (A2) implies $K = \sigma_Z^2 \sum_{i=0}^{\infty} b^{2i} < \infty$,

$$0 \leq \prod_{j=1}^i \phi_j \prod_{j=1}^{i+k} \phi_j \leq b^{2i+k},$$

and hence

$$c_2(k) = \sigma_Z^2 \sum_{i=0}^{\infty} E\left(\prod_{j=1}^i \phi_j \prod_{j=1}^{i+k} \phi_j\right) \leq K b^k.$$

□

Proof (of Theorem 3) Let

$$\xi_{t,i} \stackrel{d}{=} \prod_{j=1-i+t}^t \phi_j.$$

From Lemma 1 we have

$$\gamma_X(k) = S(k) + o(k^{2d-1})$$

where

$$S(k) = \sum_{i_1, i_2=1}^{\infty} \gamma_{\xi_{i_1, i_2}}(k)$$

and

$$\begin{aligned}\gamma_{\xi;i_1,i_2}(k) &= \text{cov}(\xi_{0,i_1}, \xi_{k,i_2}) \\ &= \text{cov}\left(\prod_{j=1-i_1}^0 \phi_j, \prod_{j=1-i_2+k}^k \phi_j\right).\end{aligned}$$

Since, by assumption (A5), $\gamma_{\xi;i_1,i_2}(k) \geq 0$, it is sufficient for a lower bound to show

$$\gamma_{\xi;i_1,i_2}(k) \sim c_{i_1,i_2} k^{2d-1} \quad (k \rightarrow \infty)$$

for at least one pair $i_1, i_2 \geq 0$. Consider for instance $i_1 = i_2 = 1$. Then

$$\gamma_{\xi;i_1,i_2}(k) = \text{cov}(\phi_0, \phi_k) = \gamma_\phi(k) \sim c_{\gamma,\phi} k^{2d-1}$$

and the result follows. \square

Proof (of Corollary 1) Let $\eta_t = \log \phi_t$. Suppose that η_0, \dots, η_m are associated for any m . Since the sum is a nondecreasing function, it follows that, for arbitrary finite index sets $I_1, I_2 \subset \mathbb{Z}$,

$$S_1 = \sum_{s \in I_1} \eta_s, S_2 = \sum_{t \in I_2} \eta_t$$

are associated. Applying nondecreasing functions preserves associativity (Esary et al. 1967). Hence,

$$T_1 = \prod_{s \in I_1} \xi_s = \exp(S_1), T_2 = \prod_{t \in I_2} \xi_t = \exp(S_2)$$

are also associated, and the result follows from Theorem 3. \square

Proof (of Theorem 4) By assumption (A7),

$$\gamma_1(k; r, r) = p(k; r, r) - p^2(r) \sim c_{\gamma,1}(r, r) k^{2d-1} \quad (k \rightarrow \infty).$$

Now,

$$E(\hat{p}_n(r)) = p(r) \cdot (1 - n^{-1}) \rightarrow p(r)$$

and

$$\begin{aligned}\text{var}(\hat{p}_n(r)) &= n^{-1} \sum_{k=-(n-2)}^{n-2} \left(1 - \frac{|k|}{n}\right) \gamma_1(k; r, r) \\ &\sim c_p(r) n^{2d-1}\end{aligned}$$

where

$$c_p(r) = \frac{c_{r,1}(r, r)}{d(2d+1)}.$$

Since X_t is equal to Z_t whenever $X_{t-1} = 0$, we have

$$\hat{\lambda}_Z - \lambda_Z = \frac{n^{-1} \sum_{t=2}^n 1\{X_{t-1} = 0\} (X_t - \lambda_Z)}{n^{-1} \sum_{t=2}^n 1\{X_{t-1} = 0\}} = \frac{n^{-1} S_{n;0}}{\hat{p}_n(0)}$$

where

$$n^{-1} S_n(0) = n^{-1} \sum_{t=2}^n 1\{X_{t-1} = 0\} (Z_t - \lambda_Z).$$

Stationarity of X_t ($t \in \mathbb{Z}$), $\text{cov}(Z_s, Z_t) = \delta_{st} \sigma_Z^2$, and independence of Z_t from X_{t-1} lead to

$$E[S_n(0)] = \sum_{t=2}^n p(0) E(Z_t - \lambda_Z) = 0$$

and

$$\begin{aligned} \text{var}(n^{-1} S_n(0)) &= n^{-2} \sum_{s,t=2}^n E[1\{X_{s-1} = 0\} 1\{X_{t-1} = 0\}] \text{cov}(Z_s, Z_t) \\ &= \frac{p(0) \sigma_Z^2}{n}. \end{aligned}$$

Furthermore,

$$\hat{\lambda}_Z - \lambda_Z = \frac{n^{-1} S_n(0)}{\hat{p}_n(0)} = \frac{n^{-1} S_n(0)}{p(0)} \cdot (1 + O_p(n^{2d-1})),$$

so that

$$\text{var}(\hat{\lambda}_Z) \sim \frac{\sigma_Z^2}{p(0)} n^{-1}.$$

Next, we consider $\hat{\phi}(r)$ for $r \geq 1$. Note that

$$\hat{\phi}(r) = \frac{1}{nr} \frac{S_n(r)}{\hat{p}_n(r)} = A_n + B_n$$

where

$$\begin{aligned} A_n &= r^{-1} \frac{n^{-1} \sum_{t=2}^n 1\{X_{t-1} = r\} (X_t - \lambda_Z)}{n^{-1} \sum_{t=2}^n 1\{X_{t-1} = r\}}, \\ B_n &= r^{-1} (\lambda_Z - \hat{\lambda}_Z). \end{aligned}$$

For A_n we have

$$\begin{aligned} A_n - \phi(r) &= \frac{1}{r} \frac{n^{-1} \sum_{t=2}^n 1\{X_{t-1} = r\} (X_t - r\phi(r) - \lambda_Z)}{n^{-1} \sum_{t=2}^n 1\{X_{t-1} = r\}} \\ &= \frac{1}{p(r)r} \frac{n^{-1} \sum_{t=2}^n 1\{X_{t-1} = r\} (X_t - r\phi(r) - \lambda_Z)}{1 + (\hat{p}(r) - p(r))/p(r)} \\ &= \frac{1}{p(r)r} M_n(r) \cdot \left(1 - \frac{\hat{p}(r) - p(r)}{p(r)} + o_p\left(n^{d-\frac{1}{2}}\right)\right) \end{aligned}$$

where

$$M_n(r) = n^{-1} \sum_{t=2}^n 1\{X_{t-1} = r\} (X_t - r\phi(r) - \lambda_Z).$$

Now, $E(M_n(r)) = 0$ and

$$\text{var}(M_n(r)) = n^{-2} \sum_{s,t=2}^n M_{n;s,t}(r)$$

where

$$M_{n;s,t}(r) = E\left[1\{X_{s-1} = r\} (X_s - r\phi(r) - \lambda_Z) \cdot 1\{X_{t-1} = r\} (X_t - r\phi(r) - \lambda_Z)\right].$$

Conditioning on $C = \{X_{s-1} = r, X_{t-1} = r, \phi_s = \phi(r), \phi_t = \phi(r)\}$, we obtain

$$\begin{aligned} M_{n;s,t}(r) &= E\left[1\{X_{s-1} = r\} (Y_s + Z_s - r\phi(r) - \lambda_Z) \mid C\right] \\ &\quad \cdot E\left[1\{X_{t-1} = r\} (Y_t + Z_t - r\phi(r) - \lambda_Z) \mid C\right] \\ &= 0, \end{aligned}$$

for $s \neq t$. Moreover, for $s = t$, we have $X_t = Y_t + Z_t$ with $Y_t \sim \text{Bin}(r, \phi(r))$ so that

$$\begin{aligned} &E\left[1\{X_{t-1} = r\} (X_t - r\phi(r) - \lambda_Z)^2 \mid C\right] \\ &= E\left[1\{X_{t-1} = r\} (Y_t - r\phi(r) + Z_t - \lambda_Z)^2 \mid C\right] \\ &= r\phi(r)(1 - \phi(r)) + \sigma_Z^2. \end{aligned}$$

Therefore,

$$\text{var}(M_n(r)) = n^{-1} p(r) [r\phi(r)(1 - \phi(r)) + \sigma_Z^2],$$

and, as $n \rightarrow \infty$,

$$\text{var}(A_n) \sim \frac{1}{p^2(r)r^2} \text{var}(M_n(r)) = C_A n^{-1}$$

with

$$C_A = \frac{r\phi(r)(1 - \phi(r)) + \sigma_Z^2}{p(r)r^2}.$$

Furthermore,

$$\text{var}(B_n) = r^{-2}\text{var}(\hat{\lambda}_Z) \sim C_B n^{-1}$$

with

$$C_B = \frac{\sigma_Z^2}{p(0)r^2}.$$

Furthermore,

$$\begin{aligned} \text{cov}(M_n, n^{-1}S_n(0)) &= n^{-2} \sum_{s,t=2}^n E[1\{X_{s-1} = r\}(X_s - r\phi(r) - \lambda_Z)1\{X_{t-1} = 0\}(Z_t - \lambda_Z)] \\ &= n^{-2} \sum_{t=2}^n E[1\{X_{t-1} = r\}1\{X_{t-1} = 0\}(X_t - r\phi(r) - \lambda_Z)(Z_t - \lambda_Z)] \\ &= 0 \end{aligned}$$

so that $\text{cov}(A_n, B_n) = o(n^{-1})$. Also note that

$$\begin{aligned} C_A + C_B &= \frac{r\phi(r)(1 - \phi(r)) + \sigma_Z^2}{p(r)r^2} + \frac{\sigma_Z^2}{p(0)r^2} \\ &= \frac{r\phi(r)(1 - \phi(r)) + \sigma_Z^2(p(0) + p(r))}{p(0)p(r)r^2}. \end{aligned}$$

Overall we obtain

$$\text{var}(\hat{\phi}(r)) \sim c_{\text{est},\phi}(r)n^{-1}$$

where

$$c_{\text{est},\phi}(r) = C_A + C_B.$$

Finally, for $r \geq 1$,

$$\begin{aligned} &\text{cov}\left(n^{-\frac{1}{2}}S_n(0), n^{-d}S_n(r)\right) \\ &= n^{-\frac{1}{2}-d} \sum_{s,t=2}^n E[1\{X_{s-1} = 0\}1\{X_{t-1} = r\}(Z_s - \lambda_Z)(X_t - \lambda_Z)] \\ &= n^{-\frac{1}{2}-d} \sum_{\substack{s,t=2 \\ s \neq t}}^n E[1\{X_{s-1} = 0\}1\{X_{t-1} = r\}]E[Z_t - \lambda_Z] \\ &= 0 \end{aligned}$$

implies

$$\lim_{n \rightarrow \infty} \text{cov} \left(\sqrt{n}(\hat{\lambda}_Z - \lambda_Z), \sqrt{n}(\hat{\phi}(r) - \phi(r)) \right) = 0.$$

□

□

Acknowledgements The second author was supported by a stipend of the Cusanuswerk, Bonn. We would like to thank the referees for their insightful comments that helped to improve the presentation of the results.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Osh, M.A., Alzaied, A.A.: First order integer-valued autoregressive (INAR (1)) process. *J. Time Ser. Anal.* **8**(3), 261–275 (1987)
- Barczy, M., Nedényi, F., Pap, G.: On aggregation of multitype Galton-Watson branching processes with immigration. *Mod. Stoch.: Theory. Appl.* (2017). <https://doi.org/10.15559/18-VMSTA95>
- Beran, J.: Statistics for long-memory processes. Chapman and Hall, CRC Press, New York (1994)
- Beran, J.: On prediction of 0-1-processes. In: Dodge, Y. (ed.) *Statistical Data Analysis based on the L1-Norm and Related Methods*, pp. 139–147. Birkhäuser, Basel/Switzerland (2002)
- Beran, J., Droullier, F.: On strongly dependent zero-inflated INAR (1) processes. *Stat. Pap.* (2023). <https://doi.org/10.1007/s00362-023-01496-z>
- Beran, J., Feng, Y., Ghosh, S., Kulik, R.: Long-memory processes-probabilistic properties and statistical methods. Springer, New York (2013)
- Bourguignon, M., Vasconcellos, L.P., Reisen, K.V.A., Ispány, M.: A Poisson INAR (1) process with a seasonal structure. *J. Stat. Comput. Simul.* **86**(2), 373–387 (2016)
- Braccini, L.: Essays in Dynamic Duration and Count Modelling. PhD Thesis, University of Pennsylvania (2015)
- Brockwell, P.J., Davis, R.A.: Time series: theory and methods. Springer, New York (1991)
- Cui, Y., Wang, Y.Y.: Estimation for random coefficient integer-valued autoregressive model under random environment. *Adv. Differ. Equ.* **2019**, 1–16 (2019)
- Daley, D.J., Rolski, T., Vesilo, R.: Long-range dependent point processes and their Palm-Khinchin distributions. *Adv. Appl. Probab.* **32**(4), 1051–1063 (2000)
- Davis, R.A., Fokianos, K., Holan, S.H., Joe, H., Livsey, J., Lund, R., Pipiras, V., Ravishanker, N.: Count time series: a methodological review. *J. Am. Stat. Assoc.* **116**(535), 1533–1547 (2021)
- Davis, R.A., Holan, S.H., Lund, R., Ravishanker, N. (eds.): *Handbook of Discrete-Valued Time Series*, 1st edn. Chapman and Hall/CRC, New York (2016)
- Dobrushin, R.L., Major, P.: Non-central limit theorems for non-linear functional of Gaussian fields. *Z. Wahrscheinlichkeitstheorie und verwandte Geb.* **50**, 27–52 (1979)
- Du, J., Li, Y.: The integer-valued autoregressive (INAR (p)) model. *J. Time Ser. Anal.* **12**, 129–142 (1991)

- Esary, J.D., Proschan, F., Walkup, D.W.: Association of random variables, with applications. *Ann. Math. Statist.* **38**(5), 1466–1474. (1967). <https://doi.org/10.1214/aoms/1177698701>
- Freeland, R.K., McCabe, B.P.M.: Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Stat. Probab. Lett.* **73**(2), 147–153 (2005)
- Gauthier, G., Latour, A.: Convergence forte des estimateurs des paramètres d'un processus GENAR(p). *Ann. des Sci. Mathématiques du Québec* **18**(1), 37–59 (1994)
- Giraitis, L., Koul, H., Surgailis, D.: Large Sample Inference For Long Memory Processes. Imperial College Press, London (2012)
- Giraitis, L., Surgailis, D.: Multivariate Appell polynomials and the central limit theorem. In: Eberlein, E., Taqqu, M.S. (eds.) *Dependence in Probability and Statistics. Progress in Probability and Statistics*, vol. 11. Birkhäuser, Boston (1986)
- Granger, C.: Long memory relationships and the aggregation of dynamic models. *J. Econom.* **14**, 227–238 (1980)
- Granger, C.W.J., Joyeux, R.: An introduction to long-range time series models and fractional differencing. *J. Time Ser. Anal.* **1**, 15–30 (1980)
- Hainaut, D., Boucher, J.-P.: Frequency and severity modelling using multifractal processes: an application to tornado occurrence in the USA and CAT Bonds. *Environ. Modeling Assess.* **19**(3), 207–220 (2014)
- Homburg, A., Weiss, C.H., Alwan, L., Frahm, G., Göb, R.: Evaluating approximate point forecasting of count processes. *Econometrics* **7**, 30 (2019)
- Hosking, J.R.M.: Fractional differencing. *Biometrika* **68**, 165–176 (1981)
- Hurvich, C.M., Deo, R., Soulier, P., Wang, Y.: Conditions for the propagation of memory parameter from durations to counts and realized volatility. *Econom. Theory* **25**(3), 764–792 (2009)
- Jentsch, C., Weiß, C.H.: Bootstrapping INAR models. *Bernoulli* **25**(3), 2359–2408 (2019)
- Jia, Y., Kechagias, S., Livsey, J., Lund, R., Pipiras, V.: Latent Gaussian count time series. *J. Am. Stat. Assoc.* **118**(541), 596–606 (2023)
- Lehmann, E.L.: Some concepts of dependence. *Ann. Math. Stat.* **37**(5), 1137–1153 (1966)
- Leonenko, N.N., Savani, V., Zhigljavsky, A.A.: Autoregressive negative binomial processes. *Ann. de l'ISUP* **51**(1–2), 25–47 (2007)
- Major, P.: Multiple Wiener-Itô Integrals with Applications to Limit Theorems. Springer, Heidelberg (1981)
- Malyshev, V.A., Minlos, R.A.: Gibbs Random Fields. Kluwer Academic, Dordrecht (1991)
- McKenzie, E.: Some simple models for discrete variate time series 1. *J. Am. Water Resour. Assoc.* **21**(4), 645–650 (1985)
- Pedeli, X., Karlis, D.: A bivariate INAR (1) process with application. *Stat. Modell.* **11**(4), 325–349 (2011)
- Pipiras, V., Taqqu, M.S.: Long-range dependence and self-similarity. Cambridge University Press, Cambridge (2017)
- Quoreshi, A.: A long-memory integer-valued time series model, INARFINA, for financial application. *Quant. Financ.* **14**(12), 2225–2235 (2014)
- Schweer, S., Weiß, C.H.: Compound Poisson INAR (1) processes: stochastic properties and testing for overdispersion. *Comput. Stat. Data Anal.* **77**, 267–284 (2014)
- Scotto, M.G., Weiß, C.H., Gouveia, S.: Thinning-based models in the analysis of integer-valued time series: a review. *Stat. Model.* **15**(6), 590–618 (2015)
- Segnon, M., Stapper, M.: Long memory conditional heteroscedasticity in count data. Tech. Rep. No. 82. University of Münster, Germany (2019)
- Silva, N., Pereira, I., Silva, M.E.: Forecasting in INAR (1) model. *REVSTAT Stat. J.* **7**(1), 119–134 (2009)
- Steutel, F.W., van Harn, K.: Discrete analogues of self-decomposability and stability. *Ann. Probab.* **7**(5), 893–899 (1979)
- Surgailis, D.: CLTs for polynomials of linear sequences: diagram formula with illustrations. In: Doukhan, P., Oppenheim, G., Taqqu, M.S. (eds.) *Theory and applications of long range dependence*, pp. 111–142. Birkhäuser, Boston, Berlin (2003)
- Taqqu, M.S.: Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Adv. Appl. Probab.* **7**(2), 249–249 (1975)
- Weiß, C.H.: An introduction to discrete-valued time series. Wiley, Hoboken (2018)
- Zheng, H., Basawa, I.V., Datta, S.: First-order random coefficient integer-valued autoregressive processes. *J. Stat. Plan. Inference* **137**(1), 212–229 (2007)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.