Mixing properties of non-stationary INGARCH(1, 1) processes

PAUL DOUKHAN[1], ANNE LEUCHT[2] and MICHAEL H. NEUMANN[3]

[1]*UMR 8088 Analyse, Géométrie et Modélisation, CY University, 2, avenue Alexandre Chauvin, 95302 Cergy-Pontoise Cedex, France. E-mail: doukhan@cyu.fr*
[2]*Research Group of Statistics and Mathematics, Universität Bamberg, Feldkirchenstraße 21, D – 96052 Bamberg, Germany. E-mail: anne.leucht@uni-bamberg.de*
[3]*Institut für Mathematik, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, D – 07743 Jena, Germany. E-mail: michael.neumann@uni-jena.de*

We derive mixing properties for a broad class of Poisson count time series satisfying a certain contraction condition. Using specific coupling techniques, we prove absolute regularity at a geometric rate not only for stationary Poisson-GARCH processes but also for models with an explosive trend. We provide easily verifiable sufficient conditions for absolute regularity for a variety of models including classical (log-)linear models. Finally, we illustrate the practical use of our results for hypothesis testing.

Keywords: Absolute regularity; coupling; INGARCH; mixing

1. Introduction

Conditional heteroscedastic processes have become quite popular for modeling the evolution of stock prices, exchange rates and interest rates. Starting with the seminal papers by Engle [12] on autoregressive conditional heteroscedastic models (ARCH) and Bollerslev [3] on generalized ARCH, numerous variants of these models have been proposed for modeling financial time series; see, for example, Francq and Zakoïan [18] for a detailed overview. More recently, integer-valued GARCH models (INGARCH) which mirror the structure of GARCH models have been proposed for modeling time series of counts; see, for example, Fokianos [14] and the recently edited volume by Davis, Holan, Lund, and Ravishanker [6].

We consider integer-valued processes where the count variable Y_t at time t, given the past, has a Poisson distribution with intensity λ_t. The intensity λ_t itself is random and it is assumed that $\lambda_t = f_t(Y_{t-1}, \lambda_{t-1}, Z_{t-1})$, for some function f_t, that is, λ_t is a function of lagged values of the count and intensity processes and a covariate Z_{t-1}. Mixing properties of such processes have been derived for a first time in Neumann [21], for a time-homogeneous transition mechanism with $\lambda_t = f(Y_{t-1}, \lambda_{t-1})$. This has been generalized by Neumann [22] to a GARCH structure of arbitrary order. In both cases a contractive condition on the intensity function f was imposed which resulted in an exponential decay of the coefficients of absolute regularity. Under a weaker semi-contractive rather than a fully contractive condition on the intensity function, Doukhan and Neumann [10] also proved absolute regularity of the count process, this time with a slower subexponential decay of the mixing coefficients. In the present paper, we extend these results in two directions. We include an exogeneous covariate process in the intensity function and we also drop the condition of time-homogeneity. This allows us to consider "weakly non-stationary" processes, see, for example, with a periodic pattern in the intensity function. Moreover, we also allow for a certain explosive behavior which could see, for example, result from a deterministic trend. As shown in the text, this requires certain modifications of the techniques used in our previous work.

In the next section, we state the precise conditions, describe our approach of deriving mixing properties, and state the main results. In Section 3, we apply these results to time-homogeneous and time-inhomogeneous linear INGARCH models, to the log-linear model proposed by Fokianos and Tjøstheim [16] as well as to mixed Poisson INGARCH models. Section 4 clarifies connections to previous work and sketches a few possible extensions. In Section 5, we discuss a possible application of our results. All proofs are deferred to a final Section 6.

2. Assumptions and main results

We derive mixing properties of an integer-valued process $\mathbf{Y} = (Y_t)_{t\in\mathbb{N}_0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where, for $t \geq 1$,

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{Pois}(\lambda_t), \tag{2.1a}$$

$$\lambda_t = f_t(Y_{t-1}, \lambda_{t-1}, Z_{t-1}), \tag{2.1b}$$

and $\mathcal{F}_s = \sigma(Y_0, \lambda_0, Z_0, \ldots, Y_s, \lambda_s, Z_s)$. Here, $\boldsymbol{\lambda} = (\lambda_t)_{t\in\mathbb{N}_0}$ is the process of random (non-negative) intensities and $\mathbf{Z} = (Z_t)_{t\in\mathbb{N}_0}$ is a sequence of $\mathbb{R}^d$-valued covariates. We assume that Z_t is independent of $\mathcal{F}_{t-1}$ and Y_t. We do not assume that the Z_t's are identically distributed since we want to include cases with a possibly unbounded trend. Note that with a slight abuse of notation and to avoid an unnecessary case-by-case analysis Pois(0) denotes the Dirac measure in 0.

In what follows we derive conditions which allow us to prove absolute regularity (β-mixing) of the process $\mathbf{X} = (X_t)_{t\in\mathbb{N}_0}$, where $X_t = (Y_t, Z_t)$. In contrast, the intensity process $(\lambda_t)_{t\in\mathbb{N}_0}$ is not mixing in general; see Remark 3 in Neumann [21] for a counterexample. We will show that, in case of a two-sided stationary process, $\lambda_t = g(X_{t-1}, X_{t-2}, \ldots)$, for a suitable function g. This allows us to conclude that the intensity process, and the joint process $((Y_t, \lambda_t, Z_t))_{t\in\mathbb{Z}}$ as well, are ergodic.

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

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

For the process $\mathbf{X} = (X_t)_{t\in\mathbb{N}_0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, the coefficients of absolute regularity at the point k are defined as

$$\beta^X(k, n) = \beta\big(\sigma(X_0, X_1, \ldots, X_k), \sigma(X_{k+n}, X_{k+n+1}, \ldots)\big)$$

and the (global) coefficients of absolute regularity as

$$\beta^X(n) = \sup\big\{\beta^X(k, n): k \in \mathbb{N}_0\big\}.$$

Our approach of proving absolute regularity is inspired by the fact that one can construct, on a suitable probability space $(\widetilde{\Omega}, \widetilde{F}, \widetilde{\mathbb{P}})$, two versions of the process $\mathbf{X}$, $(\widetilde{X}_t)_{t\in\mathbb{N}_0}$ and $(\widetilde{X}'_t)_{t\in\mathbb{N}_0}$, such that $(\widetilde{X}_0, \ldots, \widetilde{X}_k)$ and $(\widetilde{X}'_0, \ldots, \widetilde{X}'_k)$ are independent and

$$\beta^X(k, n) = \widetilde{\mathbb{P}}\big(\widetilde{X}_{k+n+r} \neq \widetilde{X}'_{k+n+r} \text{ for some } r \geq 0\big). \tag{2.2}$$

Indeed, for given $(\widetilde{X}_t)_{t\in\mathbb{N}_0}$, it follows from Berbee's lemma (see Berbee [2] or Rio [24], Lemma 5.1, for a more accessible reference) that one can construct $(\widetilde{X}'_t)_{t\geq k+n}$ following the same law as $(\widetilde{X}_t)_{t\geq k+n}$ and being independent of $(\widetilde{X}_0, \ldots, \widetilde{X}_k)$ such that (2.2) is fulfilled. Using the correct conditional distribution, we can augment $(\widetilde{X}'_t)_{t\geq k+n}$ with $\widetilde{X}'_0, \ldots, \widetilde{X}'_{k+n-1}$ such that $(\widetilde{X}'_0, \ldots, \widetilde{X}'_k)$ is independent of

$(\widetilde{X}_0, \ldots, \widetilde{X}_k)$, as required. Such an ideal coupling is usually hard to find and we do not see a chance to obtain this in the cases we have in mind. However, any coupling with $(\widetilde{X}_0, \ldots, \widetilde{X}_k)$ and $(\widetilde{X}'_0, \ldots, \widetilde{X}'_k)$ being independent provides an estimate of the mixing coefficient since then

$$\beta^X(k, n) \leq \widetilde{\mathbb{P}}(\widetilde{X}_{k+n+r} \neq \widetilde{X}'_{k+n+r} \text{ for some } r \geq 0);$$

see our arguments below.

We obtain the following estimate of the coefficients of absolute regularity at the point k.

$$\beta^X(k, n)$$
$$= \beta\big(\sigma(X_0, X_1, \ldots, X_k), \sigma(X_{k+n}, X_{k+n+1}, \ldots)\big)$$
$$\leq \beta\big(\mathcal{F}_k, \sigma(X_{k+n}, X_{k+n+1}, \ldots)\big)$$
$$= \beta\big(\sigma(\lambda_{k+1}), \sigma(X_{k+n}, X_{k+n+1}, \ldots)\big)$$
$$= \mathbb{E}\Big[\sup_{C \in \sigma(\mathcal{Z})} \big\{\big|\mathbb{P}((X_{k+n}, X_{k+n+1}, \ldots) \in C \mid \lambda_{k+1}) - \mathbb{P}((X_{k+n}, X_{k+n+1}, \ldots) \in C)\big|\big\}\Big], \quad (2.3)$$

where $\mathcal{Z} = \{A_1 \times B_1 \times \cdots \times A_m \times B_m \times \mathbb{N}_0 \times \mathbb{R}^d \times \mathbb{N}_0 \times \mathbb{R}^d \times \cdots \mid A_1, \ldots, A_m \subseteq \mathbb{N}_0, B_1, \ldots, B_m \in \mathcal{B}^d, m \in \mathbb{N}\}$ is the system of cylinder sets. Note that the last but one equality in (2.3) follows since the process $((Y_t, \lambda_t, Z_t))_{t \in \mathbb{N}_0}$ is Markovian and since the conditional distribution of (Y_t, λ_t, Z_t) under $\mathcal{F}_{t-1}$ depends only on λ_t.

Since a purely analytic approach to estimate the right-hand side of (2.3) seems to be nearly impossible, we use a *stepwise* coupling method to derive the desired result. Suppose that we have two versions of the original process $((Y_t, \lambda_t, Z_t))_{t \in \mathbb{N}_0}$, $((\widetilde{Y}_t, \widetilde{\lambda}_t, \widetilde{Z}_t))_{t \in \mathbb{N}_0}$ and $((\widetilde{Y}'_t, \widetilde{\lambda}'_t, \widetilde{Z}'_t))_{t \in \mathbb{N}_0}$, which are both defined on the same probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$. If $\widetilde{\lambda}_{k+1}$ and $\widetilde{\lambda}'_{k+1}$ are independent under $\widetilde{\mathbb{P}}$, then we obtain from (2.3) the following upper estimate of the coefficients of absolute regularity at time k:

$$\beta^X(k, n)$$
$$\leq \widetilde{\mathbb{E}}\Big[\sup_{C \in \sigma(\mathcal{Z})} \big\{\big|\widetilde{\mathbb{P}}((\widetilde{X}_{k+n}, \widetilde{X}_{k+n+1}, \ldots) \in C \mid \widetilde{\lambda}_{k+1}) - \widetilde{\mathbb{P}}((\widetilde{X}'_{k+n}, \widetilde{X}'_{k+n+1}, \ldots) \in C \mid \widetilde{\lambda}'_{k+1})\big|\big\}\Big]$$
$$\leq \widetilde{\mathbb{P}}(\widetilde{X}_{k+n+r} \neq \widetilde{X}'_{k+n+r} \text{ for some } r \in \mathbb{N}_0)$$
$$= \widetilde{\mathbb{P}}(\widetilde{X}_{k+n} \neq \widetilde{X}'_{k+n})$$
$$+ \sum_{r=1}^{\infty} \widetilde{\mathbb{P}}(\widetilde{X}_{k+n+r} \neq \widetilde{X}'_{k+n+r}, \widetilde{X}_{k+n+r-1} = \widetilde{X}'_{k+n+r-1}, \ldots, \widetilde{X}_{k+n} = \widetilde{X}'_{k+n}).$$

Thus we have just proved the following result.

Proposition 2.1. *If there are two versions, $((\widetilde{Y}_t, \widetilde{\lambda}_t, \widetilde{Z}_t))_{t \in \mathbb{N}_0}$ and $((\widetilde{Y}'_t, \widetilde{\lambda}'_t, \widetilde{Z}'_t))_{t \in \mathbb{N}_0}$, of the process $((Y_t, \lambda_t, Z_t))_{t \in \mathbb{N}_0}$ defined by (2.1a) and (2.1b) which are both defined on the same probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ such that $\widetilde{\lambda}_{k+1}$ and $\widetilde{\lambda}'_{k+1}$ are independent under $\widetilde{\mathbb{P}}$, then*

$$\beta^X(k, n) \leq \widetilde{\mathbb{P}}(\widetilde{X}_{k+n} \neq \widetilde{X}'_{k+n})$$
$$+ \sum_{r=1}^{\infty} \widetilde{\mathbb{P}}(\widetilde{X}_{k+n+r} \neq \widetilde{X}'_{k+n+r}, \widetilde{X}_{k+n+r-1} = \widetilde{X}'_{k+n+r-1}, \ldots, \widetilde{X}_{k+n} = \widetilde{X}'_{k+n}). \quad (2.4)$$

The close relationship between absolute regularity and coupling has been known for a long time. Berbee [2], Theorem 2, showed that, for two random variables X and Y defined on the same probability space, the latter one can be replaced by a random variable Y^* being independent of X and following the same distribution as Y such that the probability that Y^* differs from Y is equal to the coefficient of absolute regularity between X and Y; see also Doukhan [7], Theorem 1.2.1.1, for a more accessible reference. In our paper, we go the opposite way: Starting from a coupling result we derive an upper estimate of the coefficients of absolute regularity.

In what follows we develop a coupling strategy to keep the right-hand side of (2.4) small. To this end, we couple $\widetilde{Z}_{k+n+r}$ and $\widetilde{Z}'_{k+n+r}$ ($r \in \mathbb{N}_0$) such that they are equal with probability 1, and we apply the technique of *maximal coupling* to the count variables $\widetilde{Y}_{k+n+r}$ and $\widetilde{Y}'_{k+n+r}$. If Q_1 and Q_2 are two probability distributions on $(\mathbb{N}_0, 2^{\mathbb{N}_0})$, then one can construct random variables $\bar{X}_1$ and $\bar{X}_2$ on a suitable probability space $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{Q})$ with $\bar{Q}^{\bar{X}_i} = Q_i$, $i = 1, 2$, such that

$$\bar{Q}(\bar{X}_1 \neq \bar{X}_2) = d_{\mathrm{TV}}(Q_1, Q_2),$$

where $d_{\mathrm{TV}}(Q_1, Q_2) = \max\{|Q_1(C) - Q_2(C)| : C \subseteq \mathbb{N}_0\}$ denotes the total variation distance between Q_1 and Q_2. (An alternative representation is given by $d_{\mathrm{TV}}(Q_1, Q_2) = 1 - \sum_{k=0}^{\infty} \min\{Q_1(\{k\}), Q_2(\{k\})\}$.) In our case, we have to couple among others $\widetilde{Y}_{k+n}$ and $\widetilde{Y}'_{k+n}$. We denote by $\widetilde{\mathcal{F}}_s = \sigma(\widetilde{Y}_0, \widetilde{\lambda}_0, \widetilde{Z}_0, \widetilde{Y}'_0, \widetilde{\lambda}'_0, \widetilde{Z}'_0, \ldots, \widetilde{Y}_s, \widetilde{\lambda}_s, \widetilde{Z}_s, \widetilde{Y}'_s, \widetilde{\lambda}'_s, \widetilde{Z}'_s)$ the σ-algebra generated by all random variables up to time s. We construct $\widetilde{Y}_{k+n}$ and $\widetilde{Y}'_{k+n}$ such that, conditioned on $\widetilde{\mathcal{F}}_{k+n-1}$, they have Poisson distributions with respective intensities $\widetilde{\lambda}_{k+n}$ and $\widetilde{\lambda}'_{k+n}$ and

$$\mathbb{P}(\widetilde{Y}_{k+n} \neq \widetilde{Y}'_{k+n} \mid \widetilde{\mathcal{F}}_{k+n-1}) = d_{\mathrm{TV}}(\mathrm{Pois}(\widetilde{\lambda}_{k+n}), \mathrm{Pois}(\widetilde{\lambda}'_{k+n})).$$

Let $d : [0, \infty) \times [0, \infty) \to [0, 1]$ be any distance such that

$$d_{\mathrm{TV}}(\mathrm{Pois}(\lambda), \mathrm{Pois}(\lambda')) \leq d(\lambda, \lambda') \quad \forall \lambda, \lambda' \geq 0.$$

Examples for such distances are given by $d(\lambda, \lambda') = \sqrt{2/e}|\sqrt{\lambda} - \sqrt{\lambda'}|$ (see, e.g., Roos [25], formula (5), or Exercise 9.3.5(b) in Daley and Vere-Jones [5], page 300), and $d(\lambda, \lambda') = |\lambda - \lambda'|$. Hence, we can construct $\widetilde{Y}_{k+n}$ and $\widetilde{Y}'_{k+n}$ such that

$$\mathbb{P}(\widetilde{X}_{k+n} \neq \widetilde{X}'_{k+n} \mid \widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}) = \mathbb{P}(\widetilde{Y}_{k+n} \neq \widetilde{Y}'_{k+n} \mid \widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}) \leq d(\widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}). \tag{2.5}$$

Since Z_t is by assumption independent of $\mathcal{F}_{t-1}$ and Y_t, we choose $\widetilde{Z}_{k+n}$ and $\widetilde{Z}'_{k+n}$ such that they are equal with probability 1. In view of the other terms on the right-hand side of (2.4), we impose the following condition.

(A1) There exists some $L_1 < 1$, such that the following condition is fulfilled: If $\lambda, \lambda' \geq 0$, $Y \sim \mathrm{Pois}(\lambda)$ being independent of Z_t, then

$$\mathbb{E}[d(f_t(Y, \lambda, Z_t), f_t(Y, \lambda', Z_t))] \leq L_1 d(\lambda, \lambda') \quad \forall t \in \mathbb{N}.$$

Then, if we continue to use maximal coupling,

$$\begin{aligned}
&\mathbb{P}(\widetilde{X}_{k+n+1} \neq \widetilde{X}'_{k+n+1}, \widetilde{X}_{k+n} = \widetilde{X}'_{k+n} \mid \widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}) \\
&= \widetilde{\mathbb{E}}(\mathbb{P}(\widetilde{X}_{k+n+1} \neq \widetilde{X}'_{k+n+1} \mid \widetilde{\mathcal{F}}_{k+n}) \mathbb{1}(\widetilde{X}_{k+n} = \widetilde{X}'_{k+n}) \mid \widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}) \\
&\leq \widetilde{\mathbb{E}}(d(f_t(\widetilde{Y}_{k+n}, \widetilde{\lambda}_{k+n}, \widetilde{Z}_{k+n}), f_t(\widetilde{Y}'_{k+n}, \widetilde{\lambda}'_{k+n}, \widetilde{Z}'_{k+n})) \mathbb{1}(\widetilde{X}_{k+n} = \widetilde{X}'_{k+n}) \mid \widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n})
\end{aligned}$$

$$\leq \widetilde{\mathbb{E}}\big(d\big(f_t(\widetilde{Y}_{k+n}, \widetilde{\lambda}_{k+n}, \widetilde{Z}_{k+n}), f_t\big(\widetilde{Y}_{k+n}, \widetilde{\lambda}'_{k+n}, \widetilde{Z}_{k+n}\big)\big) \mid \widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}\big)$$

$$\leq L_1 d\big(\widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}\big).$$

Proceeding in the same way, we obtain that

$$\widetilde{\mathbb{P}}\big(\widetilde{X}_{k+n+r} \neq \widetilde{X}'_{k+n+r}, \widetilde{X}_{k+n+r-1} = \widetilde{X}'_{k+n+r-1}, \dots, \widetilde{X}_{k+n} = \widetilde{X}'_{k+n} \mid \widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}\big)$$

$$\leq L_1^r d\big(\widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}\big) \tag{2.6}$$

holds for all $r \in \mathbb{N}$. It follows from (2.4) to (2.6) that

$$\beta^X(k, n) \leq \frac{1}{1 - L_1} \widetilde{\mathbb{E}}\big[d\big(\widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}\big)\big]. \tag{2.7}$$

To proceed, we have to find an upper estimate of $\widetilde{\mathbb{E}}[d(\widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n})|]$, still under the condition that $\widetilde{\lambda}_{k+1}$ and $\widetilde{\lambda}'_{k+1}$ are independent, having the same distribution as the frequency λ_{k+1} of the original process. We make the following assumption.

(A2) There exists some $L_2 < 1$, such that the following condition is fulfilled. If $\lambda, \lambda' \geq 0$, then there exists a coupling of (Y, Z) and (Y', Z'), with $Y \sim \text{Pois}(\lambda)$, $Y' \sim \text{Pois}(\lambda')$, $Z, Z' \stackrel{d}{=} Z_t$, Z being independent of Y and Z' being independent of Y', such that

$$\mathbb{E}\big[d\big(f_t(Y, \lambda, Z), f_t\big(Y', \lambda', Z'\big)\big)\big] \leq L_2 d\big(\lambda, \lambda'\big) \quad \forall t \in \mathbb{N}.$$

If (A2) is fulfilled, we obtain that

$$\widetilde{\mathbb{E}}\big(d\big(\widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n}\big) \mid \widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1}\big) \leq L_2^{n-1} d\big(\widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1}\big).$$

Therefore, we obtain in conjunction with (2.7) that

$$\beta^X(k, n) \leq \frac{1}{1 - L_1} L_2^{n-1} \widetilde{\mathbb{E}}\big[d\big(\widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1}\big)\big]. \tag{2.8}$$

Finally, in order to obtain a good bound for $\beta^X(n)$ we have to ensure that $\sup\{\widetilde{\mathbb{E}}d(\widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1}) : k \in \mathbb{N}_0\} < \infty$. Recall that, with the above method of estimating $\beta^X(k, n)$, $\widetilde{\lambda}_{k+1}$ and $\widetilde{\lambda}'_{k+1}$ have to be independent, following the same distribution as λ_{k+1}. In the case of a stationary process, an upper bound may follow from the fact that the intensities λ_k are stochastically bounded in an appropriate sense. Such an argument, however, cannot be used if the process has an explosive behavior which means that we genuinely have to derive an upper bound for $\widetilde{\mathbb{E}}d(\widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1})$, with an appropriately chosen distance d; see the examples in the next section for the necessity of a tailor-made way of handling this problem. As above, it seems to be difficult to derive an upper bound for $\widetilde{\mathbb{E}}d(\widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1})$ in a purely analytical way. Therefore, we employ once more a coupling idea and the desired upper bound will be obtained by observing two *independent* versions $(\widetilde{\lambda}_t)_{t \in \mathbb{N}_0}$ and $(\widetilde{\lambda}'_t)_{t \in \mathbb{N}_0}$ of the original intensity process. We impose the following condition.

(A3) Let $(\widetilde{\lambda}_t)_{t \in \mathbb{N}_0}$ and $(\widetilde{\lambda}'_t)_{t \in \mathbb{N}_0}$ be two independent processes on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ which have the same distribution as $(\lambda_t)_{t \in \mathbb{N}_0}$. Suppose that there exist constants $L_3 < 1$ and $M_0, M_1 < \infty$ such that
   (i) $\widetilde{\mathbb{E}}d(\widetilde{\lambda}_0, \widetilde{\lambda}'_0) \leq M_0$,
   (ii) $\widetilde{\mathbb{E}}(d(\widetilde{\lambda}_{t+1}, \widetilde{\lambda}'_{t+1}) \mid \widetilde{\lambda}_t, \widetilde{\lambda}'_t) \leq L_3 d(\widetilde{\lambda}_t, \widetilde{\lambda}'_t) + M_1, \forall t \in \mathbb{N}.$

If (A3) is fulfilled, then

$$\widetilde{\mathbb{E}}d\big(\widetilde{\lambda}_1, \widetilde{\lambda}'_1\big) \leq \widetilde{\mathbb{E}}\big[\widetilde{\mathbb{E}}\big(d\big(\widetilde{\lambda}_1, \widetilde{\lambda}'_1\big) \mid \widetilde{\lambda}_0, \widetilde{\lambda}'_0\big)\big]$$
$$\leq L_3 \widetilde{\mathbb{E}}\big[d\big(\widetilde{\lambda}_0, \widetilde{\lambda}'_0\big)\big] + M_1$$
$$\leq L_3 M_0 + M_1.$$

Furthermore, since $((\widetilde{\lambda}_t, \widetilde{\lambda}'_t))_{t \in \mathbb{N}_0}$ is a Markov chain,

$$\widetilde{\mathbb{E}}d\big(\widetilde{\lambda}_2, \widetilde{\lambda}'_2\big) \leq \widetilde{\mathbb{E}}\big[\widetilde{\mathbb{E}}\big(\widetilde{\mathbb{E}}\big(d\big(\widetilde{\lambda}_2, \widetilde{\lambda}'_2\big) \mid \widetilde{\lambda}_1, \widetilde{\lambda}'_1\big) \mid \widetilde{\lambda}_0, \widetilde{\lambda}'_0\big)\big]$$
$$\leq M_1 + L_3\big(L_3 \widetilde{\mathbb{E}}d\big(\widetilde{\lambda}_0, \widetilde{\lambda}'_0\big) + M_1\big)$$
$$\leq M_1(1 + L_3) + L_3^2 M_0.$$

By induction we obtain that

$$\widetilde{\mathbb{E}}d\big(\widetilde{\lambda}_k, \widetilde{\lambda}'_k\big) \leq \frac{M_1}{1 - L_3} + M_0 \tag{2.9}$$

holds for all $k \in \mathbb{N}$. Now we obtain from (2.8) and (2.9) the following result.

Theorem 2.1. *Suppose that* (A1) *to* (A3) *are fulfilled.*

(i) *Then*

$$\beta^X(n) \leq L_2^{n-1} \frac{1}{1 - L_1}\left(\frac{M_1}{1 - L_3} + M_0\right).$$

(ii) *Suppose in addition that* $((Y_t, \lambda_t, Z_t))_{t \in \mathbb{Z}}$ *is a two-sided strictly stationary version of the process. Then there exists a* $(\sigma(\mathcal{Z}) - \mathcal{B})$*-measurable function* g, *where* $\mathcal{Z} = \{A_1 \times B_1 \times \cdots \times A_m \times B_m \times \mathbb{N}_0 \times \mathbb{R}^d \times \mathbb{N}_0 \times \mathbb{R}^d \times \cdots \mid A_1, \ldots, A_m \subseteq \mathbb{N}_0, B_1, \ldots, B_m \in \mathcal{B}^d, m \in \mathbb{N}\}$ *is the system of cylinder sets, such that*

$$\lambda_t = g(X_{t-1}, X_{t-2}, \ldots) \quad a.s. \tag{2.10}$$

The process $((Y_t, \lambda_t, Z_t))_{t \in \mathbb{Z}}$ *is ergodic.*

Remark 1. As it can be seen from the proofs of Corollaries 3.1 and 3.2 below, the broad applicability of this result is assured by flexibility in the choice of the metric d in (A1) to (A3); for details, see the discussion about Corollary 3.1.

In retrospect, we note that our coupling method which delivers an upper estimate for $\beta^X(n) = \sup\{\beta^X(k, n) : k \in \mathbb{N}_0\}$ consists of three phases: (2.8) shows that the upper estimate depends on the expectation of $d(\lambda, \lambda')$, where λ and λ' are independent versions of λ_{k+1}. Since this expectation can hardly be computed analytically we consider two *independent* versions, $(\widetilde{\lambda}_t)_{t \in \mathbb{N}_0}$ and $(\widetilde{\lambda}'_t)_{t \in \mathbb{N}_0}$, of the intensity process and we derive recursively an upper estimate of $\widetilde{\mathbb{E}}d(\widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1})$. Condition (A3) ensures boundedness of this expectation. Once we have a uniform bound for $\widetilde{\mathbb{E}}d(\widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1})$, we start a second coupling mechanism which keeps the probability of $\widetilde{X}_{k+n} \neq \widetilde{X}'_{k+n}$ small; see (2.4) for how

this enters the upper estimate for $\beta^X(k, n)$. This is accomplished by a coupling which leads to an exponential decay of $d(\widetilde{\lambda}_{k+n}, \widetilde{\lambda}'_{k+n})$ as $n \to \infty$; (A2) serves this purpose. And finally, it can also be seen from (2.4) that the term $\sum_{r=1}^{\infty} \widetilde{\mathbb{P}}(\widetilde{X}_{k+n+r} \neq \widetilde{X}'_{k+n+r}, \widetilde{X}_{k+n+r-1} = \widetilde{X}'_{k+n+r-1}, \ldots, \widetilde{X}_{k+n} = \widetilde{X}'_{k+n})$ contributes to the upper estimate for $\beta^X(k, n)$. For this we have to take care that $\widetilde{X}_{k+n+r}$ differs from $\widetilde{X}'_{k+n+r}$ with a small probability, given $\widetilde{X}_{k+n} = \widetilde{X}'_{k+n}, \ldots, \widetilde{X}_{k+n+r-1} = \widetilde{X}_{k+n+r-1}$. Condition (A1) is intended to keep the probability of these undesired events small.

3. Examples

3.1. Linear Poisson-INGARCH processes

In this section, we discuss some of the most popular specifications for INGARCH(1, 1) processes. We begin with a linear INGARCH(1, 1) process allowing for real-valued covariates, where

$$\lambda_{t+1} = a_t Y_t + b_t \lambda_t + Z_t. \tag{3.1}$$

Without covariates and with $a_t = a$, $b_t = b$ $\forall t$, this model has become popular for modeling count data. Rydberg and Shephard [26] proposed such a model for describing the number of trades on the New York Stock Exchange in certain time intervals and called it BIN(1, 1) model. Stationarity and other properties for this model where derived by Streett [27], Ferland et al. [13] who referred to it as INGARCH(1, 1) model, and Fokianos et al. [15]. Agosto et al. [1] generalized model (3.1) by augmenting a covariate process and coined the term PARX (Poisson autoregression with exogenous covariates). These authors also proved the existence of a stationary distribution. We study first the non-explosive case.

Corollary 3.1. *Suppose that*

   (i) *(2.1a) is fulfilled,*
   (ii) *(3.1) holds, where $a_t, b_t \geq 0$ and $L_2 = \sup\{a_t + b_t : t \in \mathbb{N}_0\} < 1$,*
   (iii) *$\sup\{\mathbb{E}Z_t : t \in \mathbb{N}_0\} < \infty$ and Z_t is a non-negative random variable (covariate) which is independent of $\lambda_0, Y_0, Z_0, \ldots, \lambda_{t-1}, Y_{t-1}, Z_{t-1}, \lambda_t, Y_t$,*
   (iv) *$\mathbb{E}\lambda_0 < \infty$.*

Then the process $(X_t)_{t \in \mathbb{N}_0}$ is absolutely regular with coefficients satisfying

$$\beta^X(n) \leq L_2^{n-1} \frac{1}{1 - L_1} M,$$

where $L_1 = \sup\{b_t : t \in \mathbb{N}_0\}$ and $M = 2(\mathbb{E}\lambda_0 + \sup\{\mathbb{E}Z_t : t \in \mathbb{N}_0\}/(1 - L_2))$.

Remark 2. As it can be seen from the proof, we obtain the same result if we consider more generally $\lambda_{t+1} = g(a_t Y_t + b_t \lambda_t + Z_t)$ for some Lipschitz function g with $\text{Lip}(g) \leq 1$ under conditions (i), (iii), and (iv) of Corollary 3.1 if $L_2 = \sup\{|a_t| + |b_t| : t \in \mathbb{N}_0\} < 1$. In particular, we obtain absolute regularity with an exponential rate for softplus INGARCH(1, 1) processes with exogenous regressors under the conditions on the coefficients a_t and b_t and on the regressors $(Z_t)_t$ mentioned above. Softplus INGARCH processes without exogenous regressors have been introduced just recently by Weiß, Zhu, and Hoshiyar [29], where $g = s_c$ is the so-called softplus function

$$s_c(x) = c \ln(1 + e^{x/c}), \quad \text{with } c > 0;$$

see also Section 4 for further details.

The proof of Corollary 3.1 relies on the application of Theorem 2.1 with the simple metric $d(\lambda, \lambda') = |\lambda - \lambda'|$. In case of an explosive INGARCH(1, 1) process, however, it could well happen that this distance is no longer appropriate. To see this, consider the simple case of a specification

$$\lambda_{t+1} = aY_t + C_t,$$

where $0 < a < 1$ and C_t being an arbitrarily large non-negative constant. Recall that our estimate (2.8) of the local coefficients of absolute regularity $\beta^X(k, n)$ contains the factor $\widetilde{\mathbb{E}}d(\widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1})$ which would be $\widetilde{\mathbb{E}}|\widetilde{\lambda}_{k+1} - \widetilde{\lambda}'_{k+1}|$ using the L_1-distance. Let $((\widetilde{Y}_t, \widetilde{\lambda}_t))_{t \in \mathbb{N}_0}$ and $((\widetilde{Y}'_t, \widetilde{\lambda}'_t))_{t \in \mathbb{N}_0}$ be two independent versions of the bivariate process. Then $\widetilde{\lambda}_{t+1} - \widetilde{\lambda}'_{t+1} = a(\widetilde{Y}_t - \widetilde{Y}'_t)$ and, conditioned on $\widetilde{\lambda}_t, \widetilde{\lambda}'_t, \widetilde{Y}_t$ and $\widetilde{Y}'_t$ are independent and Poisson distributed with respective intensities $\widetilde{\lambda}_t$ and $\widetilde{\lambda}'_t$. Since $\widetilde{\lambda}_t, \widetilde{\lambda}'_t \geq C_{t-1}$ it follows that $\widetilde{\mathbb{E}}|\widetilde{\lambda}_{t+1} - \widetilde{\lambda}'_{t+1}| \to \infty$ as $C_{t-1} \to \infty$, which means that assumption (A3) will be violated. We show that the alternative distance $|\sqrt{\lambda} - \sqrt{\lambda'}|$ saves the day. The use of such a square root transformation should not come as a big surprise. Recall that $d_{\mathrm{TV}}(\mathrm{Pois}(\lambda), \mathrm{Pois}(\lambda')) \leq d(\lambda, \lambda') \leq \sqrt{2/e}|\sqrt{\lambda} - \sqrt{\lambda'}|$. On the other hand, it is well-known that a square root transformation on Poisson variates has the effect of being variance-stabilizing. In fact, if $Y_\lambda \sim \mathrm{Pois}(\lambda)$, then $\mathbb{E}[(\sqrt{Y_\lambda} - \sqrt{\lambda})^2] \to 1/4$ as $\lambda \to \infty$; see, for example, McCullagh and Nelder [20], page 96. This transformation is similar to the Anscombe transform ($x \mapsto 2\sqrt{x + 3/8}$) which is also a classical tool to treat Poisson data. On the other hand, for small values of λ and λ', the distance $|\lambda - \lambda'|$ turns out to be more suitable when a contraction property has to be derived; see the proof of Corollary 3.2 below. In view of this, we choose

$$d(\lambda, \lambda') = \begin{cases} |\lambda - \lambda'|/M & \text{if } \sqrt{\lambda} + \sqrt{\lambda'} \leq M, \\ |\sqrt{\lambda} - \sqrt{\lambda'}| & \text{if } \sqrt{\lambda} + \sqrt{\lambda'} > M \end{cases}$$

$$= \min\{|\lambda - \lambda'|/M, |\sqrt{\lambda} - \sqrt{\lambda'}|\}, \tag{3.2}$$

where a suitable choice of the constant $M \in (0, \infty)$ becomes apparent from the proof of Corollary 3.2 below.

Corollary 3.2. *Suppose that*

$$\lambda_{t+1} = a_t Y_t + b_t \lambda_t + Z_t, \tag{3.3}$$

where

   (i) $a_t, b_t \geq 0$ *with* $\sup\{a_t + b_t : t \in \mathbb{N}_0\} < 1$,
   (ii) $\sup\{\mathbb{E}|\sqrt{Z_t} - \mathbb{E}\sqrt{Z_t}| : t \in \mathbb{N}_0\} < \infty$,
   (iii) $\mathbb{E}\sqrt{\lambda_0} < \infty$.

Then the process $(X_t)_{t \in \mathbb{N}_0}$ *is absolutely regular with coefficients satisfying*

$$\beta^X(n) = O(\rho^n)$$

for some $\rho < 1$.

Note that the random variable Z_t may get arbitrarily large as t increases, for example, it could represent a trend. Hence, we allow for nonstationary, explosive scenarios here.

3.2. Log-linear Poisson-INGARCH processes

Next, we consider the log-linear model proposed by Fokianos and Tjøstheim [16].

Proposition 3.1. *Suppose that*

$$\log(\lambda_{t+1}) = d + a \log(\lambda_t) + b \log(Y_t + 1) + Z_t, \tag{3.4}$$

where $d \in \mathbb{R}$ and $|a| + |b| < 1$, and $(Z_t)_{t \in \mathbb{N}_0}$ are i.i.d. random variables such that $\mathbb{E}|Z_0| < \infty$. Then

  (i) *there exists a (strictly) stationary version of $((Y_t, \lambda_t, Z_t))_t$,*
 (ii) *if additionally $\mathbb{E}[e^{2Z_0}] < \infty$, then the process $(X_t)_t$ is absolutely regular with exponentially decaying coefficients.*

3.3. Mixed Poisson-INGARCH processes

The above results can be generalized to models where the Poisson distribution is replaced by certain mixed Poisson distributions. We consider two cases, the zero-inflated Poisson and the negative binomial distribution, in more details. In both cases, our model can be put in the above framework by setting

$$\lambda_t = f_t(Y_{t-1}, \lambda_{t-1}, Z_{t-1}) := Z_{t-1}^{(2)} \widetilde{f}_t(Y_{t-1}, \lambda_{t-1}, Z_{t-1}^{(1)}), \tag{3.5}$$

where $Z_{t-1} = (Z_{t-1}^{(1)}, Z_{t-1}^{(2)})$ is a covariate with independent components $Z_t^{(1)}$ and $Z_t^{(2)}$, $Z_t^{(2)}$ being non-negative.

If $(Z_t^{(2)})_t$ in (3.5) is a sequence of i.i.d. $\text{Bin}(1, p)$ variables for some $p \in (0, 1)$ and if

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{Pois}(\lambda_t),$$

with $\mathcal{F}_{t-1} = \sigma(Y_0, \lambda_0, Z_0, \ldots, Y_{t-1}, \lambda_{t-1}, Z_{t-1})$ then, conditioned on $\mathcal{F}_{t-1}^{(1)} = \sigma(Y_0, \lambda_0, Z_0, \ldots, Y_{t-2}, \lambda_{t-2}, Z_{t-2}, Y_{t-1}, \lambda_{t-1}, Z_{t-1}^{(1)})$, Y_t has a zero-inflated Poisson distribution (see Lambert [19]) with parameters p and $\nu_t = \widetilde{f}_t(Y_{t-1}, \lambda_{t-1}, Z_{t-1}^{(1)})$, that is,

$$P(Y_t = k \mid \mathcal{F}_{t-1}^{(1)}) = \begin{cases} p e^{-\nu_t} \nu_t^k / k! & \text{if } k \geq 1, \\ (1 - p) + p e^{-\nu_t} & \text{if } k = 0. \end{cases}$$

Similar INGARCH models with such a distribution were considered see, for example, in Zhu [30] to account for overdispersion and potential extreme observations.

If instead $(Z_t^{(2)})_t$ has a Gamma distribution with parameters $a, b > 0$ and

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{Pois}(\lambda_t),$$

then, conditioned on $\mathcal{F}_{t-1}^{(1)}$ as above, Y_t has a negative binomial distribution. Indeed, since a Gamma(a, b) distribution has a density p with

$$p(x) = \begin{cases} \dfrac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

we obtain that, for all $k \in \mathbb{N}_0$,

$$
\begin{aligned}
P\big(Y_t = k \mid \mathcal{F}_{t-1}^{(1)}\big) &= \int_0^\infty \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} e^{-\lambda x} \frac{(\lambda x)^k}{k!} \, dx \\
&= \frac{1}{\Gamma(a)k!} b^a \lambda^k \int_0^\infty x^{a+k-1} e^{-(\lambda+b)x} \, dx \\
&= \frac{\Gamma(a+k)}{\Gamma(a)k!} \left(\frac{b}{\lambda+b} \right)^a \left(1 - \frac{b}{\lambda+b} \right)^k.
\end{aligned}
$$

This is the probability mass function of a NB$(a, b/(\lambda + b))$ distribution.

In both cases, we may use Theorem 2.1 to prove that the process $(X_t)_t$ is absolutely regular with exponentially decaying coefficients. Note that under validity of (A3), it suffices to check (A1) and (A2) for $\widetilde{f}_t$ rather than f_t (with $L_1, L_2 < b/a$ for the NB example). To see this, consider a coupling such that $\widetilde{Z}_t^{(2)} = \widetilde{Z}_t^{(2)'}$ which then gives

$$
\begin{aligned}
\widetilde{E}\big(d\big(\widetilde{Z}_t^{(2)} \widetilde{f}_t(\widetilde{Y}_t, \widetilde{\lambda}_t, \widetilde{Z}_t^{(1)}), \widetilde{Z}_t^{(2)} f\big(\widetilde{Y}_t', \widetilde{\lambda}_t', \widetilde{Z}_t^{(1)'}\big)\big) \mid \widetilde{\lambda}_t, \widetilde{\lambda}_t'\big) \\
= \widetilde{E}\big[\widetilde{Z}_t^{(2)}\big] \widetilde{E}\big(d\big(f\big(\widetilde{Y}_t, \widetilde{\lambda}_t, \widetilde{Z}_t^{(1)}\big), f\big(\widetilde{Y}_t', \widetilde{\lambda}_t', \widetilde{Z}_t^{(1)'}\big)\big) \mid \widetilde{\lambda}_t, \widetilde{\lambda}_t'\big).
\end{aligned}
$$

Those are the two most suitable cases for applications; anyway the distribution of other independent variables $Z_t^{(2)}$ for which (A1) and (A2) hold can also be considered.

4. Relation to previous work and possible perspectives

In the context of *stationary* INGARCH processes, absolute regularity with a geometric decay of the mixing coefficients of the count process has already been proved in Neumann [21] under a fully contractive condition,

$$
\big|f(y, \lambda) - f(y', \lambda')\big| \le a|y - y'| + b|\lambda - \lambda'| \quad \forall y, y' \in \mathbb{N}_0, \forall \lambda, \lambda' \ge 0, \tag{4.1}
$$

where a and b are non-negative constants with $a + b < 1$. Doukhan and Neumann [10] proved absolute regularity with a somewhat unusual subgeometric decay of the coefficients for GARCH and INGARCH processes of arbitrary order p and q under a weaker semi-contractive condition,

$$
\big|f(y_1, \ldots, y_p; \lambda_1, \ldots, \lambda_q) - f(y_1, \ldots, y_p; \lambda_1', \ldots, \lambda_q')\big| \le \sum_{i=1}^q c_i |\lambda_i - \lambda_i'| \tag{4.2}
$$

for all $y_1, \ldots, y_p \in \mathbb{N}_0$; $\lambda_1, \ldots, \lambda_q, \lambda_1', \ldots, \lambda_q' \ge 0$, where $c_1, \ldots, c_q$ are non-negative constants with $c_1 + \cdots + c_q < 1$.

For the specification (3.3) and without a covariate ($Z_t = 0 \, \forall t$), conditions (4.1) and (4.2) are both fulfilled. However, in case of a non-stationary covariate process $(Z_t)_{t \in \mathbb{N}_0}$, stationarity of the process $((Y_t, \lambda_t))_{t \in \mathbb{N}_0}$ might fail and the results in the above mentioned papers cannot be used. More seriously, in case of an explosive behavior, for example, if Z_t is non-random with $Z_t \to \infty$ as $t \to \infty$, the stability condition (2.5) in Neumann [21] as well as the drift condition (A1) in Doukhan and Neumann [10] are violated and a direct adaptation of the proofs in those papers seems to be impossible.

In case of a specification $\lambda_t = (a\sqrt{Y_{t-1}} + b\sqrt{\lambda_{t-1}})^2$ we obtain that

$$\left|\widetilde{\lambda}_{t+1} - \widetilde{\lambda}'_{t+1}\right| = \left|a^2(\widetilde{Y}_t - \widetilde{Y}'_t) + b^2(\widetilde{\lambda}_t - \widetilde{\lambda}'_t) + 2ab\left(\sqrt{\widetilde{Y}_t}\sqrt{\widetilde{\lambda}_t} - \sqrt{\widetilde{Y}'_t}\sqrt{\widetilde{\lambda}'_t}\right)\right|.$$

If $\widetilde{Y}_t$ and $\widetilde{Y}'_t$ are equal but large, then the right-hand side of this equation will be dominated by the term $2ab\sqrt{\widetilde{Y}_t}|\sqrt{\widetilde{\lambda}_t} - \sqrt{\widetilde{\lambda}'_t}|$ which shows that both (4.1) and (4.2) are violated. However, Theorem 2.1 is applicable. One can follow the lines of the proof of Corollary 3.1 to verify the validity of (A1) to (A3) for $d(\lambda, \lambda') = |\sqrt{\lambda} - \sqrt{\lambda'}|$.

We would like to mention that similar results as in our paper are possible for INGARCH models with distributions different from the Poisson. Doukhan, Mamode Khan, and Neumann [9] proved existence and uniqueness of a stationary distribution and absolute regularity of the count process for models where the Poisson distribution is replaced by the distribution of the difference of two independent Poisson variates (special case of a Skellam distribution). We expect similar results in the case of a generalized Poisson distribution which was advocated in the context of INGARCH models in Zhu [31]. Moreover, standard GARCH models with a normal distribution can be treated by this approach as well.

After our paper was completed, a referee brought to our attention a recently accepted paper by Weiß, Zhu, and Hoshiyar [29], where the log-linear function is replaced by the softplus function s_c stated in Remark 2. The corresponding Poisson-INGARCH model is specified by

$$\lambda_t = s_c\left(\alpha_0 + \sum_{i=1}^{p}\alpha_i X_{t-i} + \sum_{j=1}^{q}\beta_j\lambda_{t-j}\right).$$

These authors proved existence and uniqueness of a stationary distribution and, relying on results derived by Doukhan and Neumann [10], absolute regularity of the count process with a *sub*exponential decay rate for the corresponding mixing coefficients under weaker summability assumptions on the coefficients than in our Remark 2 (but neither allowing for exogenous regressors nor for time-varying coefficients). The setting of general observation-driven models with covariates is also considered in Doukhan, Neumann, and Truquet [11].

5. Testing for a trend in linear INARCH(1) models with application to COVID-19 data

5.1. Statistical study

Suppose that we observe $Y_0, \ldots, Y_n$ of a linear INARCH process as in (3.1) with $b_t = 0, \forall t$. We aim to test stationarity versus the presence of an isotonic trend. Thus, the null hypothesis will be that $\mathbb{E}Y_1 = \cdots = \mathbb{E}Y_n$ while the alternative can be characterized by $\mathbb{E}Y_1 \leq \mathbb{E}Y_2 \leq \cdots \leq \mathbb{E}Y_n$ with at least one strict inequality in this chain of inequalities. When we fit a linear model

$$Y_t = \theta_0 + \theta_1 t + \varepsilon_t, \quad t = 1, \ldots, n,$$

with a possibly non-stationary sequence of innovations $(\varepsilon_t)_t$, then the null hypothesis corresponds to $\theta_1 = 0$ and the alternative to $\theta_1 > 0$. (Even if the above linear model is not adequate, a projection will lead to $\theta_1 > 0$.) The following discussion will be simplified when we change over to an orthogonal regression model,

$$Y_t = \theta_0 + \theta_1 w_t + \varepsilon_t, \quad t = 1, \ldots, n,$$

where $w_t = (t - \frac{n+1}{2})/\sqrt{\sum_{s=1}^{n}(s - \frac{n+1}{2})^2}$. Then the columns in the corresponding design matrix are orthogonal and the l_2 norm of the vector composed of the entries in the second column is equal to 1. Therefore, we obtain for the least squares estimator $\widehat{\theta}_1$ of θ_1 that

$$\widehat{\theta}_1 = \sum_{t=1}^{n} w_t Y_t.$$

As before, we have $\theta_1 = \mathbb{E}\widehat{\theta}_1 > 0$ if there is any positive (linear or nonlinear) trend and $\theta_1 = 0$ under the null hypothesis. Therefore, $\widehat{\theta}_1$ can be used as a test statistic.

Proposition 5.1. *Suppose that $Y_0, \ldots, Y_n$ is a stretch of observations of a stationary* INARCH(1) *with constant coefficients such that*

$$\lambda_t = aY_{t-1} + b_0 \quad \text{with } a \in (0, 1), b_0 \geq 0.$$

Then, with $\sigma^2 = b_0/(1 - a)^3$,

$$\widehat{\theta}_1 \xrightarrow{d} Z_0 \sim \mathcal{N}(0, \sigma^2).$$

We show that the test statistic is asymptotically unbounded for the special case of a linear trend component in the intensity function. Other situations such as general polynomial trends can be treated similarly.

Proposition 5.2. *Suppose that $Y_0, \ldots, Y_n$ is a stretch of observations of a nonstationary* INARCH(1) *with constant coefficients and trend such that*

$$\lambda_t = aY_{t-1} + b_0 + b_1 t \quad \text{with } a \in (0, 1), b_0 \geq 0 \text{ and } b_1 > 0, t \in \mathbb{N}_0$$

and λ_0 has a finite absolute fourth moment. Then, for any $K > 0$

$$P(\widehat{\theta}_1 > K) \underset{n \to \infty}{\longrightarrow} 1.$$

Hence, a test rejecting the null if

$$\widehat{\theta}_1/\sigma > z_{1-\alpha}$$

is asymptotically of size α and consistent. Here, $z_{1-\alpha}$ denotes the $(1 - \alpha)$ quantile of $\mathcal{N}(0, 1)$. In practice, σ is unknown and has to be estimated consistently. For our simulations and the data example presented below, we used the corresponding OLS-estimators $\widehat{a}$ and $\widehat{b}_0$ to obtain $\widehat{\sigma}^2 = \widehat{b}_0/(1 - \widehat{a})^3$. More precisely, we considered the model stated in Proposition 5.2 and calculated

$$\begin{pmatrix} \widehat{a} \\ \widehat{b}_0 \\ \widehat{b}_1 \end{pmatrix} = (X^T X)^{-1} X^T Y \quad \text{with } Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \text{ and } X = \begin{pmatrix} Y_0 & 1 & 1 \\ \vdots & \vdots & \vdots \\ Y_{n-1} & 1 & n \end{pmatrix} \tag{5.1}$$

Lemma 5.1. *In the situation of Proposition 5.1 the OLS estimators of a and b_0 are consistent.*

With similar arguments as in the proof of Lemma 5.1 it can be shown that the OLS estimators of a and b_0 are also consistent under the alternative described in Proposition 5.2.
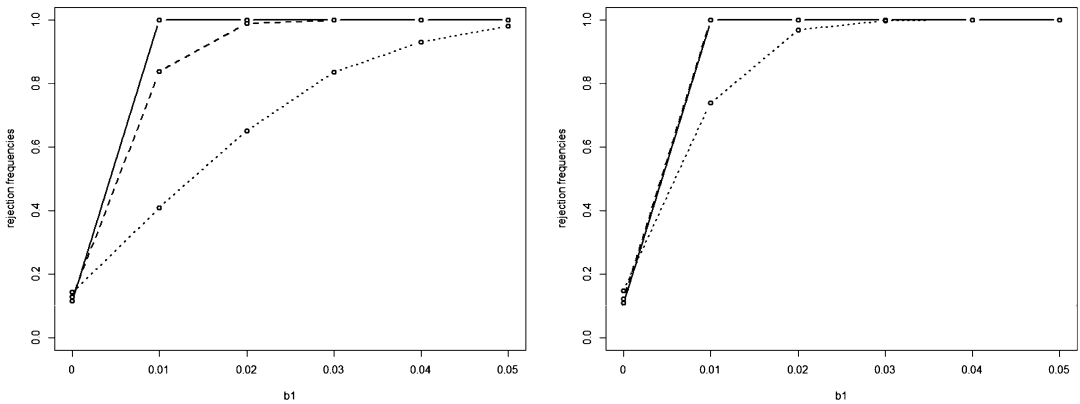
Figure 1. Left: $a = 0.2$ and $n = 50$ dotted, $n = 100$ dashed, $n = 200$ solid line; Right: $a = 0.5$ and $n = 100$ dotted, $n = 250$ dashed, $n = 500$ solid line.

Remark 3. We stick to the INARCH(1) model although a generalization of Propositions 5.1 and 5.2 to INGARCH(1, 1) models is possible. In the latter case, the naive OLS estimation is no longer feasible since the intensity process is unobserved. Of course, there are consistent estimators for stationary INGARCH(1, 1) processes as well. However, their behavior under the alternative would have to be investigated, too. This goes far beyond the scope of the paper.

5.2. Numerical study

Next, we investigate the finite sample behavior of the proposed test. Considering low and moderate levels of persistence ($a = 0.2$ and $a = 0.5$) we increase the effect of a linear trend from $b_1 = 0$ (null hypothesis) to $b_1 = 0.1$ holding the intercept fixed ($b_0 = 1$). We vary the sample size $n = 50, 100, 200$ for $a = 0.2$ and $n = 100, 250, 500$ for $a = 0.5$. The results for $\alpha = 0.1$ using 5000 Monte Carlo loops are displayed in Figure 1. The power properties of our test are very convincing however it tends to reject a true null too often in small samples. In particular, note that for $a = 0.5$ increasing the sample size from 250 to 500 improves the performance of the test under the null but barely influences the behavior under the alternative (the solid and dashed line in Figure 1 nearly coincide).

5.3. Analysis of COVID-19 data

We applied our test to investigate daily COVID-19 infection numbers as well as the cases of deaths related to COVID-19 in France and Germany from July 15 to September 15, 2020 using a data set published by the European Centre for Disease Prevention and Control [17]; see Figure 2. Observing a weekly periodicity in the data, we pre-processed the data eliminating an estimated seasonal component. Obviously, no test is required to observe an increasing trend in the daily infection numbers in France as well as in Germany. Our test clearly rejects the null in both cases (France: $\widehat{\theta}_1/\widehat{\sigma} = 349,693.30$, Germany: $\widehat{\theta}_1/\widehat{\sigma} = 96.39$). However, the situation changes if we look at the cases of deaths. Again, the null is rejected for France ($\widehat{\theta}_1/\widehat{\sigma} = 7.50$) at any reasonable level. Contrary, evaluating the test statistic based on the number of deaths in Germany that are related to COVID-19, we obtain $\widehat{\theta}_1/\widehat{\sigma} = -0.11$, that is, the null hypothesis of no trend is not rejected at any reasonable level. We also studied a shift of the window of observation of 16 days, that is, we considered the period from August 1 to September
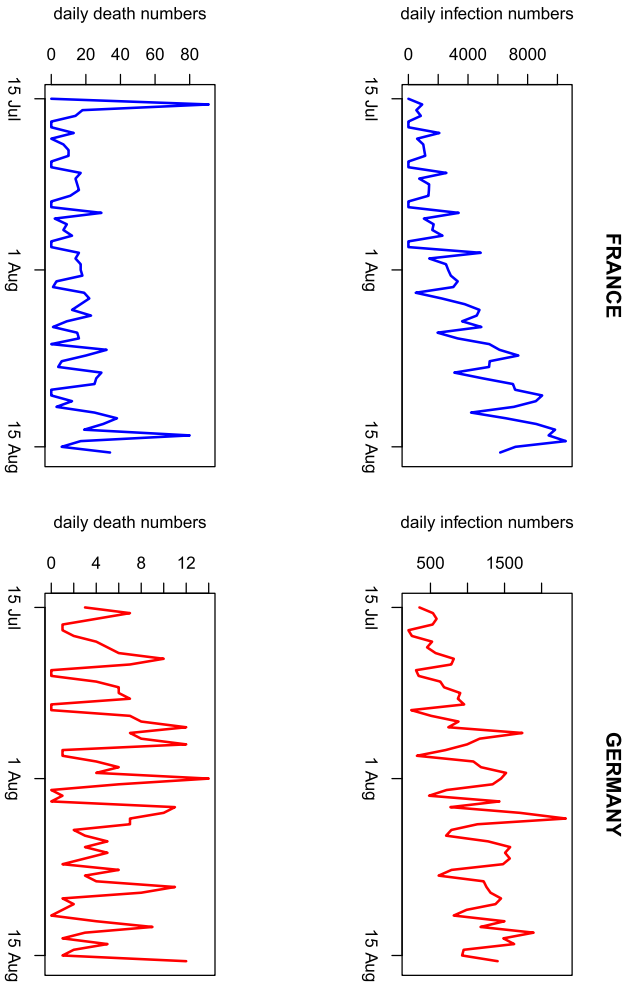
Figure 2. Daily COVID-19 infection numbers (top) as well as the cases of deaths (bottom) related to COVID-19 in France (left, blue) and Germany (right, red) from July 15 to September 15, 2020.

30. Then, unfortunately, the null is rejected for both countries for the number of daily infections as well as for the COVID-19 related number of deaths.

6. Proofs

Proof of Theorem 2.1. The proof of assertion (i) is given in the running text of Section 2.

To prove (ii), we first identify a function g, which will satisfy the required equality (2.10). We consider backward iterations $g^{[k]}$, where (with $x_i = (y_i, z_i)$) $g^{[1]}(x_1, \lambda_1) := f(y_1, \lambda_1, z_1)$ and, for $k \geq 2$, $g^{[k]}(x_1, \ldots, x_k, \lambda_k) := f(y_1, g^{[k-1]}(x_2, \ldots, x_k, \lambda_k), z_1)$. Using the idea of iterations as in exercise 46 of Doukhan [8], we consider and we set precise approximations to λ_t,

$$\lambda_t^{[k]} = g^{[k]}(X_{t-1}, \ldots, X_{t-k}, \bar{\lambda}),$$

where $\bar{\lambda} = \mathbb{E}\lambda_0$. It follows from (A2) that

$$\mathbb{E}\left[d\left(\lambda_t, \lambda_t^{[k]}\right)\right] \leq L_2^k \mathbb{E}\left[d(\lambda_{t-k}, \bar{\lambda})\right] = L_2^k \mathbb{E}\left[d(\lambda_0, \bar{\lambda})\right],$$

which implies that

$$d\left(\lambda_t, \lambda_t^{[k]}\right) \overset{\mathbb{P}}{\longrightarrow} 0 \quad \text{as } k \to \infty.$$

This implies

$$d_{\mathrm{TV}}\left(\mathrm{Pois}(\lambda_t), \mathrm{Pois}(\lambda_t^{[k]})\right) \overset{\mathbb{P}}{\longrightarrow} 0$$

and, therefore,

$$\lambda_t - \lambda_t^{[k]} \xrightarrow{\mathbb{P}} 0.$$

By taking an appropriate subsequence $(k_n)_{n \in \mathbb{N}}$ of $\mathbb{N}$ we even obtain

$$\lambda_t - \lambda_t^{[k_n]} \xrightarrow{\text{a.s.}} 0. \tag{6.1}$$

In order to obtain a well-defined function g, we define, for any sequence $x_1, x_2, \ldots,$

$$g(x_1, x_2, \ldots) = \limsup_{n \to \infty} g^{[k_n]}(x_1, \ldots, x_{k_n}, \bar{\lambda}).$$

As a limit of the measurable functions $g^{[k_n]}$, g is also $(\sigma(\mathcal{Z}) - \mathcal{B})$-measurable. From (6.1), we conclude that

$$\lambda_t = \lim_{n \to \infty} \lambda_t^{[k_n]} = g(X_{t-1}, X_{t-2}, \ldots)$$

holds with probability 1, as required.

Since absolute regularity of the process $(X_t)_{t \in \mathbb{Z}}$ implies strong mixing (see e.g. Doukhan [7], page 20) we conclude from Remark 2.6 on page 50 in combination with Proposition 2.8 on page 51 in Bradley [4] that any stationary version of this process is also ergodic. Finally, we conclude from (6.1) by Proposition 2.10(ii) in Bradley [4], page 54, that also the process $((Y_t, \lambda_t, Z_t))_{t \in \mathbb{Z}}$ is ergodic. $\qquad \square$

Proof of Corollary 3.1. We choose the distance d as $d(\lambda, \lambda') = |\lambda - \lambda'|$ and verify that conditions (A1) to (A3) are fulfilled.

(A1): We construct the coupling such that $\widetilde{Z}_t = \widetilde{Z}'_t$. Then

$$\big|\widetilde{\lambda}_{t+1} - \widetilde{\lambda}'_{t+1}\big| \mathbb{1}\big(\widetilde{Y}_t = \widetilde{Y}'_t\big) \leq b_t \big|\widetilde{\lambda}_t - \widetilde{\lambda}'_t\big|.$$

Therefore, (A1) is fulfilled with $L_1 = \sup\{b_t : t \in \mathbb{N}_0\}$.

(A2): We couple the covariates such that $\widetilde{Z}_t = \widetilde{Z}'_t$. The count variables are coupled in such a way that $\widetilde{Y}_t \geq \widetilde{Y}'_t$ if $\widetilde{\lambda}_t \geq \widetilde{\lambda}'_t$ and $\widetilde{Y}_t \leq \widetilde{Y}'_t$ if $\widetilde{\lambda}_t \leq \widetilde{\lambda}'_t$. Such a coupling is necessary and sufficient for $\widetilde{\mathbb{E}}(|\widetilde{Y}_t - \widetilde{Y}'_t| | \widetilde{\lambda}_t, \widetilde{\lambda}'_t) = |\widetilde{\lambda}_t - \widetilde{\lambda}'_t|$; otherwise the term on the left-hand side will be larger. Note that the maximal coupling but also the simple "additive coupling" share this property. The latter can be constructed as follows. If $\widetilde{\lambda}'_t \leq \widetilde{\lambda}_t$ then $\widetilde{Y}_t = \widetilde{Y}'_t + W_t$, where $W_t \sim \text{Pois}(\widetilde{\lambda}_t - \widetilde{\lambda}'_t)$ is independent of $\widetilde{Y}'_t$. Vice versa, if $\widetilde{\lambda}'_t > \widetilde{\lambda}_t$ then $\widetilde{Y}'_t = \widetilde{Y}_t + W_t$, where $W_t \sim \text{Pois}(\widetilde{\lambda}'_t - \widetilde{\lambda}_t)$ is independent of $\widetilde{Y}_t$. Then

$$\widetilde{\mathbb{E}}\big(\big|\widetilde{\lambda}_{t+1} - \widetilde{\lambda}'_{t+1}\big| \,\big|\, \widetilde{\lambda}_t, \widetilde{\lambda}'_t\big) = a_t \widetilde{\mathbb{E}}\big(\big|\widetilde{Y}_t - \widetilde{Y}'_t\big| \,\big|\, \widetilde{\lambda}_t, \widetilde{\lambda}'_t\big) + b_t \big|\widetilde{\lambda}_t - \widetilde{\lambda}'_t\big| = (a_t + b_t) \big|\widetilde{\lambda}_t - \widetilde{\lambda}'_t\big|,$$

that is, (A2) is fulfilled with $L_2 = \sup\{a_t + b_t : t \in \mathbb{N}_0\}$.

It follows from (3.1) that $\mathbb{E}\lambda_{k+1} \leq (a_k + b_k)\mathbb{E}\lambda_k + \mathbb{E}Z_k$, which implies that

$$\mathbb{E}\lambda_k \leq \mathbb{E}\lambda_0 + \frac{1}{1 - L_2} \sup\{\mathbb{E}Z_t : t \in \mathbb{N}_0\}. \qquad \square$$

Proof of Corollary 3.2. *(A1)*: We construct the coupling such that $\widetilde{Z}_t = \widetilde{Z}'_t$. Since $|\sqrt{\lambda + c} - \sqrt{\lambda' + c}| \leq |\sqrt{\lambda} - \sqrt{\lambda'}|$ holds for all $\lambda, \lambda', c \geq 0$ we obtain that

$$\big|\sqrt{\widetilde{\lambda}_{t+1}} - \sqrt{\widetilde{\lambda}'_{t+1}}\big| \mathbb{1}\big(\widetilde{Y}_t = \widetilde{Y}'_t\big) \leq \sqrt{b_t} \big|\sqrt{\widetilde{\lambda}_t} - \sqrt{\widetilde{\lambda}'_t}\big|.$$

On the other hand, the inequality $|\widetilde{\lambda}_{t+1} - \widetilde{\lambda}'_{t+1}|\mathbb{1}(\widetilde{Y}_t = \widetilde{Y}'_t) \le b|\widetilde{\lambda}_t - \widetilde{\lambda}'_t|$ is obvious. Hence, (A1) is fulfilled with $L_1 = \sup\{\sqrt{b_t} : t \in \mathbb{N}_0\}$.

(A2): We couple the covariates such that $\widetilde{Z}_t = \widetilde{Z}'_t$. For the count variables, we use an additive coupling as described in the proof of Corollary 3.1(A2). This yields in particular that $\widetilde{Y}_t \ge \widetilde{Y}'_t$ if $\widetilde{\lambda}_t \ge \widetilde{\lambda}'_t$ and $\widetilde{Y}_t \le \widetilde{Y}'_t$ if $\widetilde{\lambda}_t \le \widetilde{\lambda}'_t$. We will show that, for some $\rho < 1$,

$$\widetilde{\mathbb{E}}\big(d\big(\widetilde{\lambda}_{t+1}, \widetilde{\lambda}'_{t+1}\big) \mid \widetilde{\lambda}_t, \widetilde{\lambda}'_t\big) \le \rho d\big(\widetilde{\lambda}_t, \widetilde{\lambda}'_t\big), \tag{6.2}$$

provided that the constant M in (3.2) is chosen appropriately. To this end, we distinguish between two cases:

Case (i): $\sqrt{\widetilde{\lambda}_t} + \sqrt{\widetilde{\lambda}'_t} \le M$

Then $d(\widetilde{\lambda}_t, \widetilde{\lambda}'_t) = |\widetilde{\lambda}_t - \widetilde{\lambda}'_t|/M$ and it follows that

$$\begin{aligned}
&\widetilde{\mathbb{E}}\big(d\big(\widetilde{\lambda}_{t+1}, \widetilde{\lambda}'_{t+1}\big) \mid \widetilde{\lambda}_t, \widetilde{\lambda}'_t\big) \\
&\le \widetilde{\mathbb{E}}\big(|\widetilde{\lambda}_{t+1} - \widetilde{\lambda}'_{t+1}|/M \mid \widetilde{\lambda}_t, \widetilde{\lambda}'_t\big) \\
&= (a_t + b_t)|\widetilde{\lambda}_t - \widetilde{\lambda}'_t|/M = (a_t + b_t)d\big(\widetilde{\lambda}_t, \widetilde{\lambda}'_t\big).
\end{aligned} \tag{6.3}$$

Case (ii): $\sqrt{\widetilde{\lambda}_t} + \sqrt{\widetilde{\lambda}'_t} > M$

In this case, $d(\widetilde{\lambda}_t, \widetilde{\lambda}'_t) = |\sqrt{\widetilde{\lambda}_t} - \sqrt{\widetilde{\lambda}'_t}|$. We choose $\epsilon > 0$ such that $\sup\{\sqrt{a_t + b_t} : t \in \mathbb{N}_0\} < 1 - \epsilon$. To simplify notation, let λ, λ' be non-random with $\lambda \ge \lambda'$, $\sqrt{\lambda} + \sqrt{\lambda'} > M$ and let $Y = Y' + Z$, where $Y' \sim \text{Pois}(\lambda')$ and $Z \sim \text{Pois}(\lambda - \lambda')$ are independent. Furthermore, we drop the index t with a_t and b_t. Again, we have to distinguish between two cases.

(a): $\epsilon\sqrt{\lambda} \ge (1 + \epsilon)\sqrt{\lambda'}$

In this case the proof of (6.2) is almost trivial. We have

$$\begin{aligned}
\mathbb{E}\big[\sqrt{aY + b\lambda} &- \sqrt{aY' + b\lambda'}\big] \\
&\le \mathbb{E}\sqrt{aY + b\lambda} \\
&\le \sqrt{a + b}\sqrt{\lambda} \\
&= \sqrt{a + b}(1 + \epsilon)\left(1 - \frac{\epsilon}{1 + \epsilon}\right)\sqrt{\lambda} \\
&\le \sqrt{a + b}(1 + \epsilon)\big(\sqrt{\lambda} - \sqrt{\lambda'}\big) \\
&\le \frac{\sqrt{a + b}}{1 - \epsilon}\big|\sqrt{\lambda} - \sqrt{\lambda'}\big|.
\end{aligned} \tag{6.4}$$

Here, the second inequality follows by Jensen's inequality since $x \mapsto \sqrt{x}$ is a concave function.

(b): $\epsilon\sqrt{\lambda} < (1 + \epsilon)\sqrt{\lambda'}$

This case requires more effort. We split up

$$\begin{aligned}
&\mathbb{E}\big[\sqrt{aY + b\lambda} - \sqrt{aY' + b\lambda'}\big] \\
&\le \mathbb{E}\big[\big(\sqrt{aY + b\lambda} - \sqrt{aY' + b\lambda'}\big)\mathbb{1}\big(\sqrt{aY + b\lambda} + \sqrt{aY' + b\lambda'} \ge (1 - \epsilon)\sqrt{a + b}\big(\sqrt{\lambda} + \sqrt{\lambda'}\big)\big)\big]
\end{aligned}$$

$$+ \mathbb{E}\big[\big(\sqrt{aY + b\lambda} - \sqrt{aY' + b\lambda'}\big)\mathbb{1}\big(\sqrt{aY + b\lambda} < (1 - \epsilon)\sqrt{(a + b)\lambda} \text{ and}$$

$$\sqrt{aY' + b\lambda'} \geq (1 - \epsilon)\sqrt{(a + b)\lambda'}\big)\big]$$

$$+ \mathbb{E}\big[\big(\sqrt{aY + b\lambda} - \sqrt{aY' + b\lambda'}\big)\mathbb{1}\big(\sqrt{aY' + b\lambda'} < (1 - \epsilon)\sqrt{(a + b)\lambda'}\big)\big]$$

$$=: T_1 + T_2 + T_3, \tag{6.5}$$

say. Then

$$T_1 = \mathbb{E}\left[\frac{aY + b\lambda - aY' - b\lambda'}{\sqrt{aY + b\lambda} + \sqrt{aY' + b\lambda'}}\mathbb{1}\big(\sqrt{aY + b\lambda} + \sqrt{aY' + b\lambda'} \geq (1 - \epsilon)\sqrt{a + b}\big(\sqrt{\lambda} + \sqrt{\lambda'}\big)\big)\right]$$

$$\leq \mathbb{E}\left[\frac{aY + b\lambda - aY' - b\lambda'}{(1 - \epsilon)\sqrt{a + b}\big(\sqrt{\lambda} + \sqrt{\lambda'}\big)}\right]$$

$$= \frac{\sqrt{a + b}}{1 - \epsilon}\frac{\lambda - \lambda'}{\sqrt{\lambda} + \sqrt{\lambda'}} = \frac{\sqrt{a + b}}{1 - \epsilon}\big|\sqrt{\lambda} - \sqrt{\lambda'}\big|. \tag{6.6}$$

Since $\sqrt{aY + b\lambda} < (1 - \epsilon)\sqrt{(a + b)\lambda}$ implies that $Y < (1 - \epsilon)^2\lambda$, and therefore $|Y - \lambda| > (1 - (1 - \epsilon)^2)\lambda$, we obtain that

$$T_2 \leq \mathbb{E}\big[(1 - \epsilon)\sqrt{a + b}\big(\sqrt{\lambda} - \sqrt{\lambda'}\big)\mathbb{1}\big(|Y - \lambda| \geq \big(1 - (1 - \epsilon)^2\big)\lambda\big)\big]$$

$$\leq (1 - \epsilon)\sqrt{a + b}\big(\sqrt{\lambda} - \sqrt{\lambda'}\big)\frac{1}{(1 - (1 - \epsilon)^2)^2\lambda}$$

$$\leq \big|\sqrt{\lambda} - \sqrt{\lambda'}\big|\frac{(1 - \epsilon)\sqrt{a + b}}{(1 - (1 - \epsilon)^2)^2}\frac{4}{M^2}. \tag{6.7}$$

Note that the last inequality follows from $2\sqrt{\lambda} \geq \sqrt{\lambda} + \sqrt{\lambda'} > M$. To estimate T_3, we use the simple estimates

$$\sqrt{aY + b\lambda} - \sqrt{aY' + b\lambda'} = \sqrt{aY + b\lambda} - \sqrt{aY' + b\lambda} + \sqrt{aY' + b\lambda} - \sqrt{aY' + b\lambda'}$$

$$\leq \sqrt{a}\big(\sqrt{Y} - \sqrt{Y'}\big) + \sqrt{b}\big(\sqrt{\lambda} - \sqrt{\lambda'}\big)$$

and $\sqrt{Y} - \sqrt{Y'} \leq Y - Y'$, as well as the fact that $\sqrt{aY' + b\lambda'} < (1 - \epsilon)\sqrt{(a + b)\lambda'}$ implies that $|Y' - \lambda'| > (1 - (1 - \epsilon^2))\lambda'$. This leads to

$$T_3 \leq \sqrt{a}\mathbb{E}\big[\big(\sqrt{Y} - \sqrt{Y'}\big)\mathbb{1}\big(\sqrt{aY' + b\lambda'} < (1 - \epsilon)\sqrt{(a + b)\lambda'}\big)\big]$$

$$+ \sqrt{b}\mathbb{E}\big[\big(\sqrt{\lambda} - \sqrt{\lambda'}\big)\mathbb{1}\big(\sqrt{aY' + b\lambda'} < (1 - \epsilon)\sqrt{(a + b)\lambda'}\big)\big]$$

$$\leq \sqrt{a}\mathbb{E}\big[(Y - Y')\mathbb{1}\big(|Y' - \lambda'| > \big(1 - (1 - \epsilon^2)\big)\lambda'\big)\big]$$

$$+ \sqrt{b}\big(\sqrt{\lambda} - \sqrt{\lambda'}\big)\mathbb{P}\big(|Y' - \lambda'| > \big(1 - (1 - \epsilon^2)\big)\lambda'\big)$$

$$\leq \big(\sqrt{a}\big(\lambda - \lambda'\big) + \sqrt{b}\big(\sqrt{\lambda} - \sqrt{\lambda'}\big)\big)\frac{\mathbb{E}(Y' - \lambda')^2}{(1 - (1 - \epsilon^2))^2\lambda'^2}.$$

From $\epsilon\sqrt{\lambda} < (1 + \epsilon)\sqrt{\lambda'}$ we obtain that $M \leq \sqrt{\lambda} + \sqrt{\lambda'} \leq \frac{1 + 2\epsilon}{\epsilon}\sqrt{\lambda'}$, which leads to

$$T_3 \leq \big|\sqrt{\lambda} - \sqrt{\lambda'}\big|\frac{1}{(1 - (1 - \epsilon^2))^2}\left(\frac{1 + 2\epsilon}{\epsilon}\right)^2\left(\frac{\sqrt{a}}{M} + \frac{\sqrt{b}}{M^2}\right). \tag{6.8}$$

To sum up, we conclude from (6.3) to (6.8) that (6.2) is fulfilled for

$$\rho = \frac{\sqrt{a+b}}{1-\epsilon} + \frac{(1-\epsilon)\sqrt{a+b}}{(1-(1-\epsilon)^2)^2}\frac{4}{M^2} + \frac{1}{(1-(1-\epsilon^2))^2}\left(\frac{1+2\epsilon}{\epsilon}\right)^2\left(\frac{\sqrt{a}}{M} + \frac{\sqrt{b}}{M^2}\right).$$

Choosing now the constant M sufficiently large we obtain that $\rho < 1$, as required.

(A3): Part (i) of (A3) is fulfilled by assumption.

Assume that the processes $((\widetilde{Y}_t, \widetilde{\lambda}_t, \widetilde{Z}_t))_{t\in\mathbb{N}_0}$ and $((\widetilde{Y}'_t, \widetilde{\lambda}'_t, \widetilde{Z}'_t))_{t\in\mathbb{N}_0}$ are independent copies of the original process $((Y_t, \lambda_t, Z_t))_{t\in\mathbb{N}_0}$. We have that

$$
\begin{aligned}
&\left|\sqrt{\widetilde{\lambda}_{t+1}} - \sqrt{\widetilde{\lambda}'_{t+1}}\right| \\
&= \left|\sqrt{a\widetilde{Y}_t + b\widetilde{\lambda}_t + \widetilde{Z}_t} - \sqrt{a\widetilde{Y}'_t + b\widetilde{\lambda}'_t + \widetilde{Z}_t} + \sqrt{a\widetilde{Y}'_t + b\widetilde{\lambda}'_t + \widetilde{Z}_t} - \sqrt{a\widetilde{Y}'_t + b\widetilde{\lambda}'_t + \widetilde{Z}'_t}\right| \\
&\leq \left|\sqrt{a\widetilde{Y}_t + b\widetilde{\lambda}_t} - \sqrt{a\widetilde{Y}'_t + b\widetilde{Y}'_t}\right| + \left|\sqrt{\widetilde{Z}_t} - \sqrt{\widetilde{Z}'_t}\right| \\
&\leq \left|\sqrt{a\widetilde{Y}_t + b\widetilde{\lambda}_t} - \sqrt{(a+b)\widetilde{\lambda}_t}\right| \\
&\quad + \sqrt{a+b}\left|\sqrt{\widetilde{\lambda}_t} - \sqrt{\widetilde{\lambda}'_t}\right| \\
&\quad + \left|\sqrt{a\widetilde{Y}'_t + b\widetilde{\lambda}'_t} - \sqrt{(a+b)\widetilde{\lambda}'_t}\right| \\
&\quad + \left|\sqrt{\widetilde{Z}_t} - \sqrt{\widetilde{Z}'_t}\right| \\
&=: R_{t,1} + \cdots + R_{t,4},
\end{aligned}
\tag{6.9}
$$

say. We obtain that

$$
\begin{aligned}
\widetilde{\mathbb{E}}\left(R_{t,1}\mid\widetilde{\lambda}_t, \widetilde{\lambda}'_t\right) &= \widetilde{\mathbb{E}}\left(\frac{a|\widetilde{Y}_t - \widetilde{\lambda}_t|}{\sqrt{a\widetilde{Y}_t + b\widetilde{\lambda}_t} + \sqrt{(a+b)\widetilde{\lambda}_t}}\,\Big|\,\widetilde{\lambda}_t, \widetilde{\lambda}'_t\right) \\
&\leq \frac{a}{\sqrt{a+b}}\widetilde{\mathbb{E}}\left(|\widetilde{Y}_t - \widetilde{\lambda}_t|/\sqrt{\widetilde{\lambda}_t}\mid\widetilde{\lambda}_t, \widetilde{\lambda}'_t\right) \\
&\leq \frac{a}{\sqrt{a+b}}\sqrt{\widetilde{\mathbb{E}}\left((\widetilde{Y}_t - \widetilde{\lambda}_t)^2/\widetilde{\lambda}_t\mid\widetilde{\lambda}_t, \widetilde{\lambda}'_t\right)} = \frac{a}{\sqrt{a+b}}
\end{aligned}
\tag{6.10}
$$

and, for the same reason,

$$\widetilde{\mathbb{E}}\left(R_{t,3}\mid\widetilde{\lambda}_t, \widetilde{\lambda}'_t\right) \leq \frac{a}{\sqrt{a+b}}.\tag{6.11}$$

Finally, we have that

$$\widetilde{\mathbb{E}}\left(\left|\sqrt{\widetilde{Z}_t} - \sqrt{\widetilde{Z}'_t}\right|\mid\widetilde{\lambda}_t, \widetilde{\lambda}'_t\right) = \widetilde{\mathbb{E}}\left|\sqrt{\widetilde{Z}_t} - \sqrt{\widetilde{Z}'_t}\right| \leq 2\mathbb{E}\left|\sqrt{Z_t} - \mathbb{E}\sqrt{Z_t}\right|.\tag{6.12}$$

It follows from (6.9) to (6.12) that part (ii) of condition (A3) is fulfilled with $L_3 = \sqrt{a+b}$ and $M_0 = 2\sup\{\mathbb{E}|\sqrt{Z_t} - \mathbb{E}\sqrt{Z_t}|: t\in\mathbb{N}_0\} + 2a/\sqrt{a+b}$. $\qquad\square$

Proof of Proposition 3.1. (i) First of all, note that the process $(V_t)_{t \in \mathbb{N}_0}$ with $V_t = (\log(\lambda_t), \log(Y_t + 1), Z_t)$ forms a time-homogeneous Markov chain. Let $S = \mathbb{R} \times \log(\mathbb{N}) \times \mathbb{R}$ be the state space of this process.

In order to derive a contraction property, we choose the metric

$$\Delta\big((x, y, z), (x', y', z')\big) = \kappa_1 |x - x'| + \kappa_2 |y - y'| + |z - z'|,$$

where κ_1 and κ_2 are strictly positive constants such that $|a| \le \kappa_1$, $|b| \le \kappa_2$, and $\kappa := \kappa_1 + \kappa_2 < 1$. We show that we can couple two versions of the process $(V_t)_{t \in \mathbb{N}_0}$, $(\widetilde{V}_t)_{t \in \mathbb{N}_0}$ and $(\widetilde{V}'_t)_{t \in \mathbb{N}_0}$, such that

$$\widetilde{\mathbb{E}}\big(\Delta\big(\widetilde{V}_{t+1}, \widetilde{V}'_{t+1}\big) \mid \widetilde{V}_t, \widetilde{V}'_t\big) \le \kappa \Delta\big(\widetilde{V}_t, \widetilde{V}'_t\big). \tag{6.13}$$

We couple the corresponding covariate processes such that they coincide, i.e. $\widetilde{Z}_t = \widetilde{Z}'_t \ \forall t \in \mathbb{N}_0$. Let $v = (x, y, z)$, $v' = (x', y', z') \in S$ be arbitrary. We assume that $\widetilde{V}_t = v$ and $\widetilde{V}'_t = v'$ and construct $\widetilde{V}_{t+1} = (\log(\widetilde{\lambda}_{t+1}), \log(\widetilde{Y}_{t+1} + 1), \widetilde{Z}_{t+1})$ and $\widetilde{V}'_{t+1} = (\log(\widetilde{\lambda}'_{t+1}), \log(\widetilde{Y}'_{t+1} + 1), \widetilde{Z}'_{t+1})$ as follows. According to the model equation (3.4) we set

$$\log(\widetilde{\lambda}_{t+1}) = d + ay + bx + \widetilde{Z}_t$$

and

$$\log(\widetilde{\lambda}'_{t+1}) = d + ay' + bx' + \widetilde{Z}'_t.$$

Conditioned on $\widetilde{V}_t$ and $\widetilde{V}'_t$, the random variables $\widetilde{Y}_{t+1}$ and $\widetilde{Y}'_{t+1}$ have to follow Poisson distributions with intensities $\widetilde{\lambda}_{t+1}$ and $\widetilde{\lambda}'_{t+1}$, respectively. At this point we employ a coupling such that $\widetilde{Y}_{t+1} - \widetilde{Y}'_{t+1}$ has with probability 1 the same sign as $\widetilde{\lambda}_{t+1} - \widetilde{\lambda}'_{t+1}$. This implies in particular that

$$\begin{aligned}
&\widetilde{\mathbb{E}}\big(\big|\log(\widetilde{Y}_{t+1} + 1) - \log(\widetilde{Y}'_{t+1} + 1)\big| \mid \widetilde{V}_t, \widetilde{V}'_t\big) \\
&= \big|\widetilde{\mathbb{E}}\big(\log(\widetilde{Y}_{t+1} + 1) - \log(\widetilde{Y}'_{t+1} + 1) \mid \widetilde{V}_t, \widetilde{V}'_t\big)\big| \\
&= \big|\mathbb{E}\big(\log(Y_{t+1} + 1) \mid \lambda_{t+1} = \widetilde{\lambda}_{t+1}\big) - \mathbb{E}\big(\log(Y_{t+1} + 1) \mid \lambda_{t+1} = \widetilde{\lambda}'_{t+1}\big). \tag{6.14}
\end{aligned}$$

To estimate the term on the right-hand side of (6.14), we show that, for $Y^{(\lambda)} \sim \text{Pois}(\lambda)$,

$$\frac{d}{d\lambda}\big\{\mathbb{E}\log(Y^{(\lambda)} + 1))\big\} \le \frac{1}{\lambda} \quad \forall \lambda > 0. \tag{6.15}$$

To see this, suppose that $Y^{(\lambda)} \sim \text{Pois}(\lambda)$ and $Y^{(\epsilon)} \sim \text{Pois}(\epsilon)$ are independent. Then

$$\begin{aligned}
&\mathbb{E}\big[\log\big(Y^{(\lambda)} + Y^{(\epsilon)} + 1\big) - \log\big(Y^{(\lambda)} + 1\big)\big] \\
&= e^{-\epsilon} \epsilon \sum_{k=0}^{\infty} \big[\log(k + 2) - \log(k + 1)\big] e^{-\lambda} \frac{\lambda^k}{k!} \\
&\quad + e^{-\epsilon} \sum_{l=2}^{\infty} \frac{\epsilon^l}{l!} \sum_{k=0}^{\infty} \big[\log(k + l + 1) - \log(k + 1)\big] e^{-\lambda} \frac{\lambda^k}{k!} \\
&=: T_{\epsilon,1} + T_{\epsilon,2},
\end{aligned}$$

say. Since $\log(k+l+1) - \log(k+1) = \int_{k+1}^{k+l+1} \frac{1}{u}\,du \le \frac{l}{k+1}$ we obtain that

$$0 \le T_{\epsilon,1} = e^{-\epsilon}\epsilon \sum_{k=0}^{\infty} \frac{1}{k+1} e^{-\lambda} \frac{\lambda^k}{k!} \le \epsilon \frac{1}{\lambda} \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^{k+1}}{(k+1)!} = \frac{\epsilon}{\lambda} \mathbb{P}(Y_\lambda \ne 0)$$

as well as

$$0 \le T_{\epsilon,2} \le \frac{1}{\lambda} \sum_{l=2}^{\infty} \frac{\epsilon^\lambda}{l!} \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^{k+1}}{(k+1)!} \le \frac{1}{\lambda} \sum_{l=2}^{\infty} \epsilon^l = \frac{1}{\lambda} \frac{\epsilon^2}{1-\epsilon}.$$

Therefore,

$$\frac{d}{d\lambda} \mathbb{E}\big[\log(Y^{(\lambda)}+1)\big] = \lim_{\epsilon \to 0} \frac{T_{\epsilon,1}}{\epsilon} = \frac{\mathbb{P}(Y^{(\lambda)} \ne 0)}{\lambda} \le \frac{1}{\lambda},$$

that is, (6.15) holds true. Hence, we obtain from (6.14) that

$$\widetilde{\mathbb{E}}\big(\big|\log(\widetilde{Y}_{t+1}+1) - \log(\widetilde{Y}'_{t+1}+1)\big| \mid \widetilde{Y}_t, \widetilde{Y}'_t\big) \le \big|\log(\widetilde{\lambda}_{t+1}) - \log(\widetilde{\lambda}'_{t+1})\big|. \tag{6.16}$$

Recall that we have, by construction, $\widetilde{Z}_{t+1} = \widetilde{Z}'_{t+1}$. Using this and the above calculations we obtain

$$\begin{aligned}
&\widetilde{\mathbb{E}}\big(\Delta\big(\widetilde{V}_{t+1}, \widetilde{V}'_{t+1}\big) \mid \widetilde{V}_t, \widetilde{V}'_t\big) \\
&\quad \le \kappa_1 \big|\log(\widetilde{\lambda}_{t+1}) - \log(\widetilde{\lambda}'_{t+1})\big| + \kappa_2 \widetilde{\mathbb{E}}\big(\big|\log(\widetilde{Y}_{t+1}+1) - \log(\widetilde{Y}'_{t+1}+1)\big| \mid \widetilde{V}_t, \widetilde{V}'_t\big) \\
&\quad \le \kappa \big|\log(\widetilde{\lambda}_{t+1}) - \log(\widetilde{\lambda}'_{t+1})\big| \\
&\quad \le \kappa \big(|a| \big|\log(\widetilde{\lambda}_t) - \log(\widetilde{\lambda}'_t)\big| + |b| \big|\log(\widetilde{Y}_t+1) - \log(\widetilde{Y}'_t+1)\big|\big) \\
&\quad \le \kappa \Delta\big(\widetilde{V}_t, \widetilde{V}'_t\big).
\end{aligned} \tag{6.17}$$

It remains to translate this contraction property for random variables into a contraction property for the corresponding distributions. For the metric Δ on S, we define

$$\mathcal{P}(S) = \left\{ Q \colon Q \text{ is a probability distribution on } S \text{ with } \int \Delta(z_0, z)\,dQ(z) < \infty \right\},$$

where $z_0 \in S$ is arbitrary. For two probability measures $Q, Q' \in \mathcal{P}(S)$, we define the Kantorovich distance based on the metric Δ (also known as Wasserstein L^1 distance) by

$$\mathcal{K}\big(Q, Q'\big) := \inf_{V \sim Q, V' \sim Q'} \widetilde{\mathbb{E}} \Delta\big(V, V'\big),$$

where the infimum is taken over all random variables V and V' defined on a common probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ with respective laws Q and Q'. We denote the Markov kernel of the processes $(V_t)_{t \in \mathbb{N}_0}$ by π^V. Now we obtain immediately from (6.17) that

$$\mathcal{K}\big(Q\pi^V, Q'\pi^V\big) \le \kappa \mathcal{K}\big(Q, Q'\big). \tag{6.18}$$

The space $\mathcal{P}(S)$ equipped with the Kantorovich metric $\mathcal{K}$ is complete. Since by (6.18) the mapping π^V is contractive it follows by the Banach fixed point theorem that the Markov kernel π^V admits a unique

fixed point Q^V, that is, $Q^V \pi^V = Q^V$. In other words, Q^V is the unique stationary distribution of the process $(V_t)_{t \in \mathbb{N}_0}$. Therefore, the process $((Y_t, \lambda_t, Z_t))_{t \in \mathbb{N}_0}$ has a unique stationary distribution as well.

(ii) In this case, we do not use Theorem 2.1 to prove absolute regularity, but Proposition 2.1. To this end, we make use of a contraction property on the logarithmic scale and change over to the square root scale afterwards. As above, we construct on a suitable probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ two versions of the three-dimensional process, $((\widetilde{Y}_t, \widetilde{\lambda}_t, \widetilde{Z}_t))_{t \in \mathbb{N}_0}$ and $((\widetilde{Y}'_t, \widetilde{\lambda}'_t, \widetilde{Z}'_t))_{t \in \mathbb{N}_0}$ where these two processes evolve independently up to time k. Then $\widetilde{\lambda}_{k+1}$ and $\widetilde{\lambda}'_{k+1}$ are independent, as required. For $t = k+1, \ldots, k+n-1$, we couple these processes such that $\widetilde{Z}_t = \widetilde{Z}'_t$ as well as $\widetilde{Y}_t \geq \widetilde{Y}'_t$ if $\widetilde{\lambda}_t \geq \widetilde{\lambda}'_t$ and vice versa $\widetilde{Y}_t \leq \widetilde{Y}'_t$ if $\widetilde{\lambda}_t \leq \widetilde{\lambda}'_t$.

We obtain from (6.16) that

$$
\begin{aligned}
\widetilde{\mathbb{E}}\big(\big|\log(\widetilde{\lambda}_{t+1}) - \log(\widetilde{\lambda}'_{t+1})\big| \mid \widetilde{\lambda}_t, \widetilde{\lambda}'_t\big) \\
\leq a\big|\log(\widetilde{\lambda}_t) - \log(\widetilde{\lambda}'_t)\big| + b\widetilde{\mathbb{E}}\big(\big|\log(\widetilde{Y}_t + 1) - \log(\widetilde{Y}'_{t+1} + 1)\big| \mid \widetilde{\lambda}_t, \widetilde{\lambda}'_t\big) \\
\leq (a+b)\big|\log(\widetilde{\lambda}_t) - \log(\widetilde{\lambda}'_t)\big|
\end{aligned}
$$

holds for all $t \in \{k+1, \ldots\}$. Using this inequality $(n-1)$-times we obtain that

$$
\mathbb{E}\big(\big|\log(\widetilde{\lambda}_{k+n}) - \log(\widetilde{\lambda}'_{k+n})\big| \mid \widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1}\big) \leq \big(|a| + |b|\big)^{n-1}\big|\log(\widetilde{\lambda}_{k+1}) - \log(\widetilde{\lambda}'_{k+1})\big|. \tag{6.19}
$$

For $t = k+n, k+n+1, \ldots$, we use a maximal coupling of the count variables, that is,

$$
\widetilde{\mathbb{P}}\big(\widetilde{Y}_t \neq \widetilde{Y}'_t \mid \widetilde{\lambda}_t, \widetilde{\lambda}'_t\big) = d_{\mathrm{TV}}\big(\mathrm{Pois}(\widetilde{\lambda}_t), \mathrm{Pois}(\widetilde{\lambda}'_t)\big).
$$

This implies by Proposition 2.1 that

$$
\begin{aligned}
\beta^X(k,n) \\
= \widetilde{\mathbb{P}}\big(\widetilde{Y}_{k+n} \neq \widetilde{Y}'_{k+n} \mid \widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1}\big) \\
+ \sum_{r=1}^{\infty} \widetilde{\mathbb{P}}\big(\widetilde{Y}_{k+n+r} \neq \widetilde{Y}'_{k+n+r}, \widetilde{Y}_{k+n+r-1} = \widetilde{Y}'_{k+n+r-1}, \ldots, \widetilde{Y}_{k+n} = \widetilde{Y}'_{k+n} \mid \widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1}\big) \\
= \sum_{r=0}^{\infty} \widetilde{\mathbb{E}}\big(d_{\mathrm{TV}}\big(\mathrm{Pois}(\widetilde{\lambda}_{k+n+r}), \mathrm{Pois}(\widetilde{\lambda}'_{k+n+r})\big) \mid \widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1}\big). \tag{6.20}
\end{aligned}
$$

Finally, it remains to make the transition from our estimates of $|\log(\widetilde{\lambda}_t) - \log(\widetilde{\lambda}'_t)|$ to the above total variation distances. Since $x \mapsto e^{x/2}$ is a convex function we have, for $0 \leq x \leq y$, $|e^{x/2} - e^{y/2}| = \int_{x/2}^{y/2} e^{u/2}/2\, du \leq \frac{e^{x/2} + e^{y/2}}{8}|x - y|$, which implies that

$$
\big|\sqrt{\widetilde{\lambda}_{k+n+r}} - \sqrt{\widetilde{\lambda}'_{k+n+r}}\big| \leq \frac{\sqrt{\widetilde{\lambda}_{k+n+r}} + \sqrt{\widetilde{\lambda}'_{k+n+r}}}{8}\big|\log(\widetilde{\lambda}_{k+n+r}) - \log(\widetilde{\lambda}'_{k+n+r})\big|. \tag{6.21}
$$

Using this and the estimate $d_{\mathrm{TV}}(\mathrm{Pois}(\lambda), \mathrm{Pois}(\lambda')) \leq \sqrt{2/e}|\sqrt{\lambda} - \sqrt{\lambda'}|$ we obtain

$$
\begin{aligned}
\widetilde{\mathbb{P}}\big(\widetilde{Y}_{k+n} \neq \widetilde{Y}'_{k+n}\big) = \widetilde{\mathbb{E}}\big[d_{\mathrm{TV}}\big(\mathrm{Pois}(\widetilde{\lambda}_{k+n}), \mathrm{Pois}(\widetilde{\lambda}'_{k+n})\big)\big] \\
\leq \sqrt{\frac{2}{e}}\widetilde{\mathbb{E}}\big[\big|\sqrt{\widetilde{\lambda}_{k+n}} - \sqrt{\widetilde{\lambda}'_{k+n}}\big|\big]
\end{aligned}
$$

$$\leq \sqrt{\frac{2}{e}} \sqrt{\widetilde{\mathbb{E}}\big(\big(\sqrt{\widetilde{\lambda}_{k+n}} + \sqrt{\widetilde{\lambda}'_{k+n}}\big)/8\big)^2} \sqrt{\widetilde{\mathbb{E}}\big(\log(\widetilde{\lambda}_{k+n}) - \log(\widetilde{\lambda}'_{k+n})\big)^2}$$

$$\leq \sqrt{\frac{1}{2e}} \sqrt{\mathbb{E}[\lambda_{k+n}^2]} \big(|a| + |b|\big)^{n-1} \sqrt{\widetilde{\mathbb{E}}\big(\log(\widetilde{\lambda}_{k+1}) - \log(\widetilde{\lambda}'_{k+1})\big)^2} \qquad (6.22)$$

and, analogously,

$$\widetilde{\mathbb{P}}\big(\widetilde{Y}_{k+n+r} \neq \widetilde{Y}'_{k+n+r}, \widetilde{Y}_{k+n+r-1} = \widetilde{Y}'_{k+n+r-1}, \ldots, \widetilde{Y}_{k+n} = \widetilde{Y}'_{k+n} \mid \widetilde{\lambda}_{k+1}, \widetilde{\lambda}'_{k+1}\big)$$

$$\leq \sqrt{\frac{1}{2e}} \sqrt{\mathbb{E}[\lambda_{k+n}^2]} \big(|a| + |b|\big)^{n-1} |a|^l \sqrt{\widetilde{\mathbb{E}}\big(\log(\widetilde{\lambda}_{k+1}) - \log(\widetilde{\lambda}'_{k+1})\big)^2}. \qquad (6.23)$$

It remains to show that $\mathbb{E}[\lambda_{k+n}^2]$ is bounded. If $Y \sim \text{Pois}(\lambda)$, then $E[(Y+1)^2] = \lambda^2 + 3\lambda + 1$. This implies

$$\mathbb{E}\big(\lambda_{t+1}^2 \mid \lambda_t\big) = e^{2d} \mathbb{E}\big[e^{2Z_0}\big] \lambda_t^{2a} \big[(\lambda_t + 2)(\lambda_t + 1)\big]^b \leq C_1 \big(\lambda_t^{2(a+b)} + 1\big),$$

for some $C_1 < \infty$. Therefore we obtain that

$$\mathbb{E}\big(\lambda_{t+1}^2 \mid \lambda_t\big) \leq C_0 \lambda_t^2 + C_2,$$

for appropriate $C_0 < 1$ and $C_2 < \infty$. From this recursion, we conclude that $\mathbb{E}[\lambda_{k+n}^2]$ is bounded. (6.22) and (6.23) yield that

$$\sup\big\{\beta^X(k,n) \colon k \in \mathbb{N}_0\big\} = O\left(\big(|a| + |b|\big)^{n-1} \sum_{r=0}^{\infty} |a|^r\right) = O\big(\big(|a| + |b|\big)^n\big). \qquad \square$$

Proof of Proposition 5.1. First, note that the contraction condition $a \in (0,1)$ assures existence of a strictly stationary version of the process with β-mixing coefficients tending to zero at a geometric rate (see Corollary 3.1 and Theorem 2.1 in Neumann [21]). (Alternatively, since we are in the stationary case, Theorem 3.1 in Neumann [21] containing both results.) Moreover, all moments of Y_t are finite, see, for example, Weiß [28], Example 4.1.6. Asymptotic normality of $\widehat{\theta}_1$ can be deduced from Application 1 in Rio [23] setting $a_{i,n} = w_i$ and $\xi_i = Y_i - EY_i$ if $\sigma^2 = \lim_{n\to\infty} \text{var}(\widehat{\theta}_1) > 0$. To this end, note that from $\sum_{t=1}^n w_t = 0$ and stationarity, we get

$$\widehat{\theta}_1 = \sum_{t=1}^n w_t (Y_t - EY_t).$$

Additionally, straight-forward calculations yield

$$s_n := \frac{1}{n^3} \sum_{t=1}^n \left(t - \frac{n+1}{2}\right)^2 = \frac{1}{12} + o(1).$$

From Weiß [28], Example 4.1.6, we know that

$$\text{cov}(Y_0, Y_h) = a^h \frac{b_0}{(1-a)^2(1+a)}$$

which gives

$$\sigma^2 \cdot \frac{(1-a)^2(1+a)}{b_0}$$

$$= 1 + \lim_{n\to\infty} \frac{2}{s_n n^3} \sum_{t=2}^{n} \left(t - \frac{n+1}{2} \right) a^t \sum_{s=1}^{t-1} \left(s - \frac{n+1}{2} \right) a^{-s}$$

$$= 1 + \lim_{n\to\infty} \frac{2}{s_n n^3} \sum_{t=2}^{n} \left(t - \frac{n+1}{2} \right) a^t \left[\frac{(t-1)a^{-(t+1)} - ta^{-t} + a^{-1}}{(a^{-1}-1)^2} - \frac{n+1}{2} \frac{a^{-t} - a^{-1}}{a^{-1} - 1} \right]$$

$$= 1 + \lim_{n\to\infty} \frac{2a}{s_n n^3 (1-a)} \sum_{t=2}^{n} \left(t - \frac{n+1}{2} \right)^2$$

$$= 1 + \frac{2a}{1-a}$$

and finally yields the desired result. □

Proof of Proposition 5.2. We split up

$$\widehat{\theta_1} = \sum_{t=1}^{n} w_t (Y_t - EY_t) + \sum_{t=1}^{n} w_t EY_t. \tag{6.24}$$

First, note that the second sum tends to infinity. To see this, rewrite

$$EY_t = aEY_{t-1} + b_0 + b_1 t = \cdots = a^t EY_0 + \sum_{k=0}^{t-1} a^k \big(b_0 + b_1 (t-k) \big).$$

As $\sqrt{\sum_{s=1}^{n} (s - \frac{n+1}{2})^2} \geq C_1 n^{3/2}$, we obtain $\sup |w_t| \leq C_2 n^{-1/2}$ which implies

$$\sum_{t=1}^{n} w_t EY_t = o(n) + b_1 \sum_{t=1}^{n} w_t \sum_{k=0}^{t-1} a^k t$$

$$= o(n) + b_1 \sum_{t=1}^{n} w_t t \frac{a^t - 1}{a - 1}$$

$$= o(n) + \frac{b}{1-a} \sum_{t=1}^{n} t w_t$$

$$= C_3 n^{3/2} + o\big(n^{3/2}\big),$$

for some positive, finite constants C_1, C_2, C_3. It remains to show that the first sum in (6.24) is $o_P(n^{3/2})$. To this end, we consider

$$E\left[\sum_{t=1}^{n} w_t (Y_t - EY_t) \right]^2 \leq \frac{1}{n} \sum_{h=-(n-1)}^{n-1} \sqrt{\beta^X(|h|)} \sum_{s=\max\{1,1-h\}}^{\min\{n,n-h\}} \sqrt[4]{E(Y_{s+h} - EY_{s+h})^4 E(Y_s - EY_s)^4}$$

applying the covariance inequality for α-mixing processes in Doukhan [7], Theorem 3(1), or Theorem 1.1 in Rio [24], the fact that the α-mixing coefficients can be bounded from above by the corresponding β-mixing coefficients and Corollary 3.2. Recall that the 2nd and the 3rd central moment of a Pois(λ) distributed random variable is just λ while the fourth central moment is $\lambda^2 + 3\lambda$. Using the binomial theorem and $E\lambda_s^2 = O(s^2)$, we can further bound

$$
\begin{aligned}
E&(Y_s - EY_s)^4 \\
&= E[\lambda_s^2 + 3\lambda_s] + 4E[\lambda_s(\lambda_s - EY_s)] + 6E[\lambda_s(\lambda_s - EY_s)^2] + E(\lambda_s - EY_s)^4 \\
&= E[\lambda_s^2 + 3\lambda_s] + 4aE[\lambda_s(Y_{s-1} - EY_{s-1})] + 6a^2 E[\lambda_s(Y_{s-1} - EY_{s-1})^2] + a^4 E(Y_{s-1} - EY_{s-1})^4 \\
&= E[\lambda_s^2 + 3\lambda_s] + 4a^2 E(Y_{s-1} - EY_{s-1})^2 + 6a^2 E[\lambda_s(Y_{s-1} - EY_{s-1})^2] + a^4 E(Y_{s-1} - EY_{s-1})^4 \\
&= E[\lambda_s^2 + 3\lambda_s] + a^2[4 + 6E\lambda_s]E(Y_{s-1} - EY_{s-1})^2 + 6a^3 E(Y_{s-1} - EY_{s-1})^3 \\
&\quad + a^4 E(Y_{s-1} - EY_{s-1})^4 \\
&\leq \widetilde{C}s^2 + a^2 \widetilde{C}s E(Y_{s-1} - EY_{s-1})^2 + 6a^3 E(Y_{s-1} - EY_{s-1})^3 + a^4 E(Y_{s-1} - EY_{s-1})^4 \\
&\leq \bar{C}s^2 + 6a^3 E(Y_{s-1} - EY_{s-1})^3 + a^4 E(Y_{s-1} - EY_{s-1})^4 \\
&\leq \bar{C}s^2 + 6a^3 [C's^2 + a E(Y_{s-2} - EY_{s-2})^3] + a^4 E(Y_{s-1} - EY_{s-1})^4 \\
&\leq C''s^2 + a^4 E(Y_{s-1} - EY_{s-1})^4.
\end{aligned}
$$

Iterating these calculations yields that $E(Y_s - EY_s)^4 = O(s^2)$ which concludes the proof. $\square$

Proof of Lemma 5.1. Rewrite $Y_t = aY_{t-1} + b_0 + \eta_t$ with $\eta_t = Y_t - \lambda_t$, $t = 1, \dots, n$. Using the corresponding matrix notation and the definition of X, we have to show that $(X^T X)^{-1} X^T \eta = o_P(1)$, where $\eta = (\eta_1, \dots, \eta_n)^T$. We proceed in two steps. First, we show that $N X^T \eta = o_P(1)$ with $N = \mathrm{diag}(n^{-1}, n^{-1}, n^{-2})$. Second, we show that $(N X^T X)^{-1} = O_P(1)$.

For the first part, straight forward calculations show that

$$
N X^T \eta = \frac{1}{n} \sum_{t=1}^{n} \begin{pmatrix} Y_{t-1}(Y_t - \lambda_t) \\ Y_t - \lambda_t \\ (Y_t - \lambda_t)t/n \end{pmatrix} = o_P(1).
$$

For the second part, we rewrite $(N X^T X)^{-1} = M(N X^T X M)^{-1}$ with $M = \mathrm{diag}(1, 1, n^{-1})$ and show that $N X^T X M$ converges stochastically to an invertible matrix. To this end, note that

$$
N X^T X M = \frac{1}{n} \begin{pmatrix} \sum\limits_{t=0}^{n-1} Y_t^2 & \sum\limits_{t=0}^{n-1} Y_t & n^{-1}\sum\limits_{t=0}^{n-1} tY_t \\[2mm] \sum\limits_{t=0}^{n-1} Y_t & n & (n+1)/2 \\[2mm] n^{-1}\sum\limits_{t=0}^{n-1} tY_t & (n+1)/2 & (n+1)(2n+1)/(6n) \end{pmatrix}
$$

$$
= \begin{pmatrix} EY_0^2 & EY_0 & EY_0/2 \\ EY_0 & 1 & 1/2 \\ EY_0/2 & 1/2 & 1/3 \end{pmatrix} + o_P(1)
$$

due to the exponentially decaying autocovariance function of $(Y_t)_t$. Finally, straight forward calculations show that the determinant of the remaining matrix is positive which concludes the proof. □

Acknowledgments

We thank two anonymous referees for their valuable comments that led to a significant improvement of the paper.

Funding

References

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

[2] Berbee, H.C.P. (1979). *Random Walks with Stationary Increments and Renewal Theory*. *Mathematical Centre Tracts* **112**. Amsterdam: Mathematisch Centrum. MR0547109

[3] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. MR0853051 https://doi.org/10.1016/0304-4076(86)90063-1

[4] Bradley, R.C. (2007). *Introduction to Strong Mixing Conditions*. *Vol*. 1. Heber City, UT: Kendrick Press. MR2325294

[5] Daley, D.J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. *Springer Series in Statistics*. New York: Springer. MR0950166

[6] Davis, R.A., Holan, S.H., Lund, R. and Ravishanker, N., eds. (2016). *Handbook of Discrete-Valued Time Series*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. Boca Raton, FL: CRC Press. MR3642975

[7] Doukhan, P. (1994). *Mixing*: *Properties and Examples*. *Lecture Notes in Statistics* **85**. New York: Springer. MR1312160 https://doi.org/10.1007/978-1-4612-2642-0

[8] Doukhan, P. (2018). *Stochastic Models for Time Series*. *Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*] **80**. Cham: Springer. MR3852404 https://doi.org/10.1007/978-3-319-76938-7

[9] Doukhan, P., Khan, N.M. and Neumann, M.H. (2021). Mixing properties of integer-valued GARCH processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **18** 401–420. MR4219669 https://doi.org/10.30757/alea.v18-18

[10] Doukhan, P. and Neumann, M.H. (2019). Absolute regularity of semi-contractive GARCH-type processes. *J. Appl. Probab.* **56** 91–115. MR3981148 https://doi.org/10.1017/jpr.2019.8

[11] Doukhan, P., Neumann, M.H. and Truquet, L. (2020). Stationarity and ergodic properties for some observation-driven models in random environments. Preprint. Available at arXiv:2007.07623.

[12] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121 https://doi.org/10.2307/1912773

[13] Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process. *J. Time Series Anal.* **27** 923–942. MR2328548 https://doi.org/10.1111/j.1467-9892.2006.00496.x

[14] Fokianos, K. (2012). Count time series. In *Time Series Analysis*: *Methods and Applications* (T. Subba Rao, S. Subba Rao and C.R. Rao, eds.). *Handbook of Statistics* **30** 315–347. Amsterdam: Elsevier/North-Holland.

[15] Fokianos, K., Rahbek, A. and Tjøstheim, D. (2009). Nonlinear Poisson autoregression. *J. Amer. Statist. Assoc.* **104** 1430–1439. MR2596998 https://doi.org/10.1198/jasa.2009.tm08270

[16] Fokianos, K. and Tjøstheim, D. (2011). Log-linear Poisson autoregression. *J. Multivariate Anal*. **102** 563–578. MR2755016 https://doi.org/10.1016/j.jmva.2010.11.002

[17] European Centre for Disease Prevention and Control (2020-10-02). Available at https://data.europa.eu/euodp/de/data/dataset/covid-19-coronavirus-data.

[18] Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models*: *Structure*, *Statistical Inference and Financial Applications*. Chichester: Wiley. MR3186556 https://doi.org/10.1002/9780470670057

[19] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34** 1–14.

[20] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. *Monographs on Statistics and Applied Probability*. London: CRC Press. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

[21] Neumann, M.H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **17** 1268–1284. MR2854772 https://doi.org/10.3150/10-BEJ313

[22] Neumann, M.H. (2021). Bootstrap for integer-valued GARCH(p, q) processes. *Stat. Neerl*. Published online 19 February 2021. Available at https://onlinelibrary.wiley.com/doi/epdf/10.1111/stan.12238.

[23] Rio, E. (1995/97). About the Lindeberg method for strongly mixing sequences. *ESAIM Probab. Stat*. **1** 35–61. MR1382517 https://doi.org/10.1051/ps:1997102

[24] Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. *Probability Theory and Stochastic Modelling* **80**. Berlin: Springer. MR3642873 https://doi.org/10.1007/978-3-662-54323-8

[25] Roos, B. (2003). Improvements in the Poisson approximation of mixed Poisson distributions. *J. Statist. Plann. Inference* **113** 467–483. MR1965122 https://doi.org/10.1016/S0378-3758(02)00095-2

[26] Rydberg, T.H. and Shephard, P. (2000). A modeling framework for the prices and times of trades made on the New York Stock Exchange. In *Nonlinear and Nonstationary Signal Processing* (W.J. Fitzgerald, R.L. Smith, A.T. Walden and P.C. Young, eds.) 217–246. Cambridge: Cambridge Univ. Press.

[27] Streett, S.B. (2000). Some observation driven models for time series of counts. Ph.D. Thesis, Colorado State Univ., Department of Statistics. MR2701787

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

[29] Weiß, C.H., Zhu, F. and Hoshiyar, A. (2022). Softplus INGARCH model. *Statist. Sinica* **32**. https://doi.org/10.5705/ss.202020.0353

[30] Zhu, F. (2011). A negative binomial integer-valued GARCH model. *J. Time Series Anal*. **32** 54–67. MR2790672 https://doi.org/10.1111/j.1467-9892.2010.00684.x

[31] Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *J. Math. Anal. Appl*. **389** 58–71. MR2876481 https://doi.org/10.1016/j.jmaa.2011.11.042