Abstract
We consider integer-valued GARCH processes, where the count variable conditioned on past values of the count and state variables follows a so-called Skellam distribution. Using arguments for contractive Markov chains we prove that the process has a unique stationary regime. Furthermore, we show asymptotic regularity ($\beta$-mixing) with geometrically decaying coefficients for the count process. These probabilistic results are complemented by a statistical analysis, a few simulations as well as an application to recent COVID-19 data.
1. Introduction and Notation

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

\[
X_t \mid \mathcal{F}_{t-1} \sim \text{Skellam}(\lambda_t),
\]

\[
\lambda_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \lambda_{t-j},
\]

and \(\mathcal{F}_s = \sigma(X_s, \lambda_s, X_{s-1}, \lambda_{s-1}, \ldots)\) denotes the \(\sigma\)-field generated by the random variables up to time \(s\). Skellam(\(\lambda\)) denotes the distribution of the difference between two independent random variables following a Poisson distribution with parameter \(\lambda\). This distribution was first investigated by Irwin (1937) and is a special case of a Skellam distribution with both parameters equal to \(\lambda\); see Skellam (1946).

Alomani et al. (2018) considered such a Skellam-GARCH process of order \(p = q = 1\) and derive the estimating equations for a conditional maximum likelihood estimator of the parameters. However, perhaps because of the absence of suitable probabilistic tools for such models, they did not provide a further analysis of the asymptotic properties of this estimator. These authors also provided an overview of related results and applied the model to differences of non-negative data of counts of monthly drug crimes.

In our contribution, we primarily focus on stochastic properties of Skellam processes of general order \(p\) and \(q\), which are indispensable for a deeper analysis of statistical procedures. We assume that the parameters \(\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q\) in the above model are non-negative and that

\[
2 \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j =: L < 1. \tag{1.2}
\]

We show in Section 2 that this yields a contraction property for a related Markov chain, \((Z_t)_t\), where \(Z_t = (X_t^2, \ldots, X_{t-p+1}^2, \lambda_t, \ldots, \lambda_{t-q+1})\) in terms of a suitable Wasserstein metric but Remark 2.1 points out the fact that contraction in the mean do not seem to hold. Using arguments found in Eberle (2019, Chapter 3) and Douc et al. (2018, Theorem 20.3.4) we show that this implies the existence and uniqueness of a stationary regime of \((Z_t)_t\), and therefore of \(((X_t, \lambda_t))_t\), too. Furthermore, we use the contraction property once more to prove absolute regularity (\(\beta\)-mixing) with exponentially decaying coefficients of the count process \((X_t)_t\).

We are convinced that these results can serve as a basis for further work with these and related models. As an example, population dynamics will be considered after differentiation, in order to rate the speed or the acceleration of the evolution of species under consideration; indeed both characteristics may be either positive or negative.

As an illustration of their usefulness, we apply in Section 3 our results to prove asymptotic normality of a least squares estimator of the parameters of a Skellam-ARCH model. This is complemented by simulations and by an analysis of recent COVID-19 data from Mauritius. We think that the approach used in this paper can also be used for many models with a similar structure. For example, we could replace the Skellam distribution by the distribution of the difference of independent variables following a mixed Poisson distribution. As mentioned in Christou and Fokianos (2014), this includes the cases with negative binomial and zero-inflated Poisson distributions which are of interest if data show underdispersion. All proofs and a few auxiliary results are collected in a final Section 4.
2. Properties of the process

2.1. Contractivity. First of all, note that the processes $Y = (Y_t)_{t \in \mathbb{Z}}$ and $Z = (Z_t)_{t \in \mathbb{Z}}$ with $Y_t = (X_t, \ldots, X_{t-p+1}, \lambda_t, \ldots, \lambda_{t-q+1})$ and $Z_t = (X^2_t, \ldots, X^2_{t-p+1}, \lambda_t, \ldots, \lambda_{t-q+1})$ are time-homogeneous Markov chains. In the following we derive a contraction property of $(Z_t)_{t \in \mathbb{Z}}$ in terms of a suitable Wasserstein metric. Inspired by [Eberle (2019, Chapter 3) and Douc et al. (2018) Theorem 20.3.4] we use this to derive existence and uniqueness of a stationary distribution of this process. Furthermore, we also show that we can exploit the contraction property of the process $(Z_t)_{t}$ to derive almost effortlessly absolute regularity of the count process. As a starting point, we state a simple coupling result for two Skellam variables.

**Lemma 2.1.** Let $\lambda, \lambda' > 0$ be arbitrary and let $U \sim \text{Unif}([0,1])$. Then

$$X := F_{\lambda}^{-1}(U) \sim \text{Skellam}(\lambda) \quad \text{and} \quad X' := F_{\lambda'}^{-1}(U) \sim \text{Skellam}(\lambda')$$

where $F_{\lambda}$ and $F_{\lambda'}$ are the respective cumulative distribution functions of Skellam distributions with parameters $\lambda$ and $\lambda'$. ($G^{-1}(t) = \inf \{x: G(x) \geq t\}$ denotes the generalized inverse of a generic distribution function $G$.)

Then the random variables $X$ and $X'$ have the same sign and $E[X^2 - X'^2] = 2|\lambda - \lambda'|$.

We use this result to derive a contraction property of the process $(Z_t)_{t \in \mathbb{Z}}$. Let $S = \{0,1^2,2^2,3^2,\ldots\}^p \times [0,\infty)^p$ be the state space of this process and let $\mathcal{P}(S \times S)$ be the family of all probability distributions supported in $S \times S$.

To formulate a coupling result, we use a metric $\Delta_{\gamma,\delta}$ on $S$ which is defined by

$$\Delta_{\gamma,\delta}((y_1, \ldots, y_p, \lambda_1, \ldots, \lambda_q), (y'_1, \ldots, y'_p, \lambda'_1, \ldots, \lambda'_q)) = \sum_{i=1}^{p} \gamma_i |y_i - y'_i| + \sum_{j=1}^{q} \delta_j |\lambda_j - \lambda'_j|,$$

where $\gamma_1, \ldots, \gamma_p, \lambda_1, \ldots, \lambda_q$ are strictly positive constants. Let now $z = (x_1^2, \ldots, x_p^2, \lambda_1, \ldots, \lambda_q), \quad z' = (x'_1^2, \ldots, x'_p^2, \lambda'_1, \ldots, \lambda'_q) \in S$ and let $U \sim \text{Unif}([0,1])$. According to Lemma 2.1 we define random vectors

$$Z = (X^2, x_1^2, \ldots, x_{p-1}^2, \lambda, \lambda_1, \ldots, \lambda_{q-1}), \quad \text{and} \quad Z' = (X'^2, x'_1^2, \ldots, x'_{p-1}^2, \lambda', \lambda'_1, \ldots, \lambda'_{q-1}),$$

where

$$\lambda = \omega + \sum_{i=1}^{p} \alpha_i x_i^2 + \sum_{j=1}^{q} \beta_j \lambda_j, \quad \lambda' = \omega + \sum_{i=1}^{p} \alpha_i x'_i^2 + \sum_{j=1}^{q} \beta_j \lambda'_j$$

and $X = F_{\lambda}^{-1}(U), \quad X' = F_{\lambda'}^{-1}(U)$. We denote the corresponding Markov kernel by $\tilde{\pi}$, that is, for the above random variables $Z$ and $Z'$ we have that $(Z, Z') \sim \tilde{\pi}((z, z'), \cdot)$.

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

**Proposition 2.1.** Suppose that condition (I.2) is fulfilled. Then there exist strictly positive constants $\gamma_1, \ldots, \gamma_p, \delta_1, \ldots, \delta_q$ such that the following contraction properties hold true.

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

$$Z \sim \mathbb{P}[Z_t | Z_{t-1} = z] \quad \text{and} \quad Z' \sim \mathbb{P}[Z_t | Z_{t-1} = z']$$

and

$$\mathbb{E} \Delta_{\gamma,\delta}(Z, Z') \leq \kappa \Delta_{\gamma,\delta}(z, z').$$
(ii) Let \( ((\tilde{Z}_t, \widetilde{Z}_t))_{t \in \mathbb{Z}} \) be a Markov chain with transition kernel \( \bar{\pi} \). Then
\[
\mathbb{E} \Delta_{\gamma, \delta} (\tilde{Z}_t, \widetilde{Z}_t) \leq \kappa \mathbb{E} \Delta_{\gamma, \delta} (\tilde{Z}_{t-1}, \widetilde{Z}_{t-1}) .
\]

2.2. Stationarity. In order to derive stationarity properties of the process \((Y_t)_{t \in \mathbb{Z}}\), we first translate the contraction result in Proposition 2.1 to a contraction property of the corresponding distributions. For two probability measures \( Q \) and \( Q' \) on \( S \), we define the Kantorovich distance based on the metric \( \Delta_{\gamma, \delta} \) (also known as Wasserstein \( L^1 \) distance) by
\[
\mathcal{K}_{\gamma, \delta}(Q, Q') := \inf_{Z \sim Q, Z' \sim Q'} \mathbb{E} \Delta_{\gamma, \delta}(Z, Z'),
\]
where the infimum is taken over all random variables \( Z \) and \( Z' \) defined on a common probability space with respective laws \( Q \) and \( Q' \). We denote the Markov kernel of the process \((Z_t)_{t \in \mathbb{Z}}\) by \( \pi \). The following result follows immediately from Proposition 2.1.

**Proposition 2.2.** Let \( Q \) and \( Q' \) be arbitrary distributions on the state space \( S \). Then, for \( \kappa < 1 \) given in Proposition 2.1,
\[
\mathcal{K}_{\gamma, \delta}(\pi Q, \pi Q') \leq \kappa \mathcal{K}_{\gamma, \delta}(Q, Q') .
\]

Proposition 2.2 shows that the mapping \( \pi \) is contractive. Therefore, we can conclude by the Banach fixed point theorem that the Markov process \((Z_t)_{t \in \mathbb{Z}}\) and consequently \((Y_t)_{t \in \mathbb{Z}}\) have unique stationary distributions.

**Theorem 2.1.** The Markov process \((Y_t)_{t \in \mathbb{Z}}\) has a unique stationary distribution \( Q \).

**Remark 2.1.** The reader might wonder why we focus on absolute regularity rather than weak dependence properties introduced more recently by Doukhan and Louhichi (1999). The following reasons led us to resort to this more classical notion. For \( X \sim \text{Skellam}(\lambda) \), we have that \( EX = 0 \) and \( EX^2 = 2\lambda \). Therefore it is natural to model the intensity process as in (1.1b), where the intensities enter linearly while the count variables are squared. Therefore, and since \( X_t \) conditioned on \( F_{t-1} \) follows a centered law, we were not able to apply arguments for the contraction in mean as in Doukhan and Louhichi (1999) in order to show existence and uniqueness of a stationary distribution, as done for example in Doukhan and Winterberger (2008). The mismatch of powers in (1.1b) is the reason why we establish in Proposition 2.1 a contraction property w.r.t. a distance where the intensities and the count variables enter with the appropriate different powers. Using this, it seems to be also possible to derive an alternative property of weak dependence, called \( \tau \)-dependence.

For sake of completeness we recall some facts from the monograph Dedecker et al. (2007). Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space, \( \mathcal{M} \) a \( \sigma \)-algebra and \( X \) be a random variable with values in \( \mathbb{R}^d \). We assume that \( \mathbb{E}\|X\| < \infty \) and define the coefficient \( \tau \) as
\[
\tau(\mathcal{M}, X) = \sup \left\{ \left\| \int f(x) \mathbb{P}_X|_{\mathcal{M}} (dx) - \int f(x) \mathbb{P}_X (dx) \right\|_1 \right\} \text{ Lip } f \leq 1 \quad (2.1)
\]
for \( f: \mathbb{R}^d \to \mathbb{R}, \) 1-Lipschitz. Then \( \tau(\mathcal{M}, X) \leq \mathbb{E}\|X - X^*\| \) if \( X^* \) is distributed as \( X \) and independent of \( \mathcal{M} \). It provides a simple way to calculate \( \tau \) and Proposition 3.2 in Dedecker et al. (2007) proves that \( L^1 \)-contraction implies a geometric decay of the coefficients \( \tau(r) \) for Markov chains. Consider an \( \mathbb{R}^d \) valued stationary time series \((X_t)_{t \in \mathbb{Z}}\). The definition
of $\tau$ allows to evaluate the dependence between the past of the sequence $(X_t)_{t\in\mathbb{Z}}$ and its $k$-tuples in the future. With the norm $\|x - y\| = \|x_1 - y_1\| + \cdots + \|x_k - y_k\|$ on $(\mathbb{R}^d)^k$, define

$$\tau(r) = \sup_{p,t \geq 1} \frac{1}{t} \sup \{ \tau(\sigma(X_t, t \leq p), (X_{j_1}, \ldots, X_{j_i})), p + r \leq j_1 < \cdots < j_i \}.$$ 

The time series $(X_t)_{t\in\mathbb{Z}}$ is said to be $\tau$-weakly dependent if $\lim_{r \to \infty} \tau(r) = 0$. The coefficient $\theta \leq \tau$ is more simply described in terms of the covariances of functions

$$\coV(g(X_{i_1}, \ldots, X_{i_u}), f(X_{j_1}, \ldots, X_{i_v}))$$

for 1-Lipschitz functions $f, \|g\|_{\infty} \leq 1$ and $i_1 < \cdots < i_u \leq p$. The usefulness of those concepts in statistics was also explained in [Doukhan and Neumann (2008)] where it was shown that many tools known to hold in the case of mixing random variables have their counterparts under such weak dependence conditions.

[Doukhan, Fokianos and and Tjostheim (2012)] derive the weak dependence properties of Poisson-based GARCH-type models with $\lambda_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}$. The original properties of Skellam models make also natural the inhomogeneity of (1.1a) and (1.1b) which include both linear and squared factors. In the simplest case of a SkellamARCH(1)-process, we write $X_t = S_t(\lambda_t)$ where $(S_t)$ denotes an i.i.d. sequence of Skellam processes (typically $S = P - P'$ for 2 independent standard Poisson processes) and $\lambda_t = \omega + \alpha_1 X_{t-1}^2$. Thus $X_t = F(X_{t-1}, S_t)$ with $F(x, S) = S(\omega + \alpha_1 x^2)$ and $L^1$-contraction arguments write

$$E[F(x', S) - F(x, S)] \leq E[P(\omega + \alpha_1 x'^2) - P(\omega + \alpha_1 x^2)] + E[P'(\omega + \alpha_1 x'^2) - P'(\omega + \alpha_1 x^2)] = 2\alpha_1|x'^2 - x^2|.$$

Thus contraction does not hold even under the condition $2\alpha_1 < 1$ required to get second order moments for the stationary solution of the model in [Lemma 4.1]. This means that Proposition 3.2 in [Dedecker et al. (2007)] does not help to conclude $\tau$-dependence in this case and we better use another power argument to derive weak dependence. Based on the symmetry of the distribution of $S_t(\lambda)$ this is easy to prove that $\sigma_t = \text{sign}(X_t)$ and $|X_t|$ are independent; moreover $P(\sigma_t = \pm 1) = 1/2$. Then the process $(Y_t)_t$ (with $Y_t = X_t^2$) is again a Markov chain $Y_t = G(Y_{t-1}, S_t)$ with $G(y, S) = S^2(\omega + \alpha_1 y)$ and now, for $y' > y$, we use the fact that $P$ and then $S$ admit independent increments, and the monotonicity of the function $\lambda \mapsto S^2(\lambda)$ implies:

$$E[G(y', S) - G(y, S)] = E[S^2(\omega + \alpha_1 y') - S^2(\omega + \alpha_1 y)]$$

$$= E(S(\omega + \alpha_1 y') - S(\omega + \alpha_1 y))^2$$

$$+ 2ES(\omega + \alpha_1 y)E(S(\omega + \alpha_1 y') - S(\omega + \alpha_1 y))$$

$$= 2\alpha_1(y' - y) + 0.$$

This provides a way to prove $\tau$-dependence of the process $(Y_t)$ as in [Doukhan and Winterberger (2008)]. Now set $X_t = \sigma_t\sqrt{Y_t}$, on the set $(Y_t \neq 0)$, the function $u \in \mathbb{N} \mapsto \sqrt{u} \in \mathbb{N}$ is $\frac{1}{2}$-Lipschitz and heredity properties yield geometric $\tau$-dependence of $((\tilde{X}_t)_t$ with $\tilde{X}_t = X_t\mathbb{1}_{\{X_t \neq 0\}}$, see [Dedecker et al. (2007)], by mimicking (2.2) in the present situation. Hence $\tau$-dependence of $((|X_t|)_t$ follows since the variables $X_t, \tilde{X}_t$ coincide on the events $(X_t \neq 0)$ and $(X_t = 0)$. The independence of $\sigma_t$ and $|X_t|$, yield the $\tau$-dependence properties $(X_t)_t$.

To conclude this overview let us quote that heredity of the weak dependence notions naturally hold through Lipschitz functions $(h(X_t))_t$ inherits dependence properties of $(X_t)_t$, and strong mixing properties are hereditary those measurable functions. Those points led
us to consider mixing conditions rather than the above weak dependence conditions to derive asymptotic theory for the statistical analysis.

2.3. Absolute regularity. For the related case of Poisson count processes with a GARCH-type structure, absolute regularity has been first proved for contractive INGARCH(1,1) processes in Neumann (2011). This has been generalized in Doukhan and Neumann (2019) to semi-contractive models and in Doukhan et al. (2020) to the case of possibly non-stationary processes. In all of these papers, the mixing properties were derived by an explicit coupling of two versions of the processes which were tailor-made for the respective properties of the processes. In this paper, our approach is slightly different. We derive both stationarity and mixing properties on the basis of a one-step contractivity property given in Proposition 2.1.

Let \((\Omega, \mathcal{A}, P)\) be a probability space and \(A_1, A_2\) be two sub-\(\sigma\)-algebras of \(\mathcal{A}\). Then the coefficient of absolute regularity is defined as

\[
\beta(A_1, A_2) = E \left[ \sup \{ |P(B \mid A_1) - P(B)| : B \in A_2 \} \right].
\]

For the process \(X = (X_t)_{t \in \mathbb{Z}}\) on \((\Omega, \mathcal{F}, P)\), the coefficients of absolute regularity are defined as

\[
\beta^X(n) = \beta(\sigma(X_0, X_{-1}, \ldots), \sigma(X_n, X_{n+1}, \ldots)).
\]

We obtain the following estimate.

\[
\begin{align*}
\beta^X(n) & \leq \beta(\mathcal{F}_0, \sigma(X_n, X_{n+1}, \ldots)) \\
& = E \left[ \sup_{C \in \sigma(\mathcal{Z})} \left\{ |P((X_n, X_{n+1}, \ldots) \in C \mid \mathcal{F}_0) - P((X_n, X_{n+1}, \ldots) \in C)| \right\} \right].
\end{align*}
\]

where \(\mathcal{Z} = \{A_1 \times \cdots \times A_m \times \mathbb{Z} \times \mathbb{Z} \times \cdots | A_1, \ldots, A_m \in \mathbb{Z}, m \in \mathbb{N}\}\) is the system of cylinder sets.

At this point we employ a coupling argument. Let \(((\tilde{Y}_t, \tilde{Y}'_t))_{t \in \mathbb{N}_0}\) be a Markov chain on a probability space \((\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})\) with transition kernel \(\tilde{P}\) and independent initial variables \(\tilde{Y}_0, \tilde{Y}'_0 \sim Q\). Then

\[
\begin{align*}
E \left[ \sup_{C \in \sigma(\mathcal{Z})} \left\{ |P((X_n, X_{n+1}, \ldots) \in C \mid \mathcal{F}_0) - P((X_n, X_{n+1}, \ldots) \in C)| \right\} \right] & \leq \tilde{P} \left( \tilde{X}_{n+k} \neq \tilde{X}'_{n+k} \right) \quad \text{for some } k \geq 0 \\
& \leq \sum_{k=0}^{\infty} \tilde{P} \left( \tilde{X}_{n+k} \neq \tilde{X}'_{n+k} \right).
\end{align*}
\]

Since \(\tilde{X}_{n+k}\) and \(\tilde{X}'_{n+k}\) are integer valued and have the same sign we obtain that

\[
\tilde{P} \left( \tilde{X}_{n+k} \neq \tilde{X}'_{n+k} \right) \leq \frac{1}{\gamma_1} \tilde{E} \Delta_{\gamma, \delta}(\tilde{Z}_{n+k}, \tilde{Z}'_{n+k}).
\]

Furthermore, by Proposition 2.1

\[
\tilde{E} \Delta_{\gamma, \delta}(\tilde{Z}_{n+k}, \tilde{Z}'_{n+k}) \leq \kappa^{n+k} \tilde{E} \Delta_{\gamma, \delta}(\tilde{Z}_0, \tilde{Z}'_0),
\]

where \(\tilde{E} \Delta_{\gamma, \delta}(\tilde{Z}_0, \tilde{Z}'_0) < \infty\). From (2.3) to (2.6) we obtain absolute regularity of the count process \((X_t)_{t \in \mathbb{Z}}\) with exponentially decaying coefficients.

**Theorem 2.2.** Suppose that condition \((1.2)\) is fulfilled and that the process \((Y_t)_{t \in \mathbb{Z}}\) is stationary. Then there exists some \(\rho < 1\) such that

\[
\beta^X(n) = O(\rho^n).
\]
3. Applications

We choose to develop the asymptotic theory for the OLSE of Skellam models (3.1) as the most standard application of the above results. Much more may be done including tests of goodness-of-fit as in [Doukhan et al. (2020)]. Prediction or model selection issues are also important and should be developed theoretically. Additional numerical issues include a simulation study of this OLSE and a partial study of the COVID-19 data with a prediction study (see Remark 3.1). Additional research work will make use of the bound of absolute regularity for many other questions such a more quantitative study of prediction, qualitative tests of goodness-of-fit such as model choice problems, or more non parametric based statistics or resampling or subsampling procedures.

3.1. OLSE of a Skellam-ARCH model. We consider the special case of an Skellam-ARCH($p$) model, where (1.1b) reduces to

$$\lambda_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2. \quad (3.1)$$

We assume that $\omega > 0$, and that $\alpha_1, \ldots, \alpha_p$ are non-negative with $\alpha = \sum_{i=1}^{p} \alpha_i < 1/\sqrt{12}$. We further assume that the process $((X_t, \lambda_t))_{t \in \mathbb{Z}}$ is in its unique stationary regime. On the basis of observations $X_1, \ldots, X_n$, we intend to estimate the vector of unknown parameters $\theta = (\omega, \alpha_1, \ldots, \alpha_p)^T$. We embed the observed random variables into a linear regression model,

$$X_t^2 = 2 \omega + \sum_{i=1}^{p} 2 X_{t-i}^2 \alpha_i + \varepsilon_t, \quad t = p + 1, \ldots, n,$$

where $\varepsilon_t = X_t^2 - 2\lambda_t$ satisfies $E(\varepsilon_t | F_{t-1}) = 0$ a.s. Then the ordinary least squares estimator is given by

$$\widehat{\theta}_n \in \arg \min_{\theta} \sum_{t=p+1}^{n} (X_t^2 - 2 (\omega + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2))^2$$

$$= \arg \min_{\theta} \|Y(n) - X(n)\theta\|^2,$$

where

$$Y(n) = \begin{pmatrix} X_{n+1}^2 \\ \vdots \\ X_n^2 \end{pmatrix}, \quad X(n) = 2 \begin{pmatrix} 1 & X_p^2 & \cdots & X_1^2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n-1}^2 & \cdots & X_{n-p}^2 \end{pmatrix}.$$ 

If the matrix $X_T(n) X(n)$ is regular, then $\widehat{\theta}_n$ is uniquely defined and

$$\widehat{\theta}_n = \left( X_T(n) X(n) \right)^{-1} X_T(n) Y(n), \quad (3.2)$$

which implies that

$$\sqrt{n} \left( \widehat{\theta}_n - \theta \right) = \left( \frac{1}{n} X_T(n) X(n) \right)^{-1} \frac{1}{\sqrt{n}} X_T(n) \varepsilon(n), \quad (3.3)$$

where $\varepsilon(n) = (\varepsilon_{p+1}, \ldots, \varepsilon_n)^T$.

The condition $\sum_{i=1}^{p} \alpha_i < 1/\sqrt{12}$ ensures by Lemma 4.1 that $EX_n^4 < \infty$. Hence, we obtain from the ergodic theorem that

$$\frac{1}{n} X_T(n) X(n) \overset{a.s.}{\to} \Sigma = 4 \begin{pmatrix} 1 & EX_1^2 & \cdots & EX_p^2 \\ EX_1^2 & EX_1^2 X_1^2 & \cdots & EX_1^2 X_p^2 \\ \vdots & \vdots & \ddots & \vdots \\ EX_p^2 & EX_p^2 X_1^2 & \cdots & EX_p^2 X_p^2 \end{pmatrix}. \quad (3.4)$$
Lemma 4.2 below shows that $\Sigma$ is a regular matrix which means that equation (3.3) holds true with a probability tending to 1. Furthermore, it follows from a central limit theorem for sums of martingale differences that

$$W_n := \frac{1}{\sqrt{n}} X^T(n) \varepsilon(n) \xrightarrow{d} Z_0,$$

where $Z_0 \sim \mathcal{N}(0_{p+1}, \eta^2 \Sigma)$, where $\eta^2 = \mathrm{E}\varepsilon^2_t$. Actually, by the Cramér-Wold device, it suffices to show that

$$c^T W_n \xrightarrow{d} c^T Z_0 \sim \mathcal{N}(0, \eta^2 c^T \Sigma c) \quad \forall c \in \mathbb{R}^{p+1}. \tag{3.5}$$

Let $c = (c_0, \ldots, c_p)^T \in \mathbb{R}^{p+1}$ be arbitrary and let

$$W_{nt} = \frac{1}{\sqrt{n}} (c_0 + \sum_{i=1}^{p} c_i X_{t-i+1}^2) \varepsilon_t, \quad t > p.$$

Then

$$c^T W_n = \sum_{t=p+1}^{n} W_{nt}.$$

According to Corollary 3.1 in Hall and Heyde (1980, page 58), (3.6) follows from

$$E(W_{nt} | \mathcal{F}_{t-1}) = 0 \quad \text{a.s.}, \tag{3.7a}$$

$$\sum_{t=p+1}^{n} E(W_{nt}^2 | \mathcal{F}_{t-1}) \xrightarrow{P} \eta^2 c^T \Sigma c, \tag{3.7b}$$

and the conditional Lindeberg condition

$$L_n(\delta) := \sum_{t=p+1}^{n} E(W_{nt}^2 1(|W_{nt}| \geq \delta) | \mathcal{F}_{t-1}) \xrightarrow{P} 0 \quad \forall \delta > 0. \tag{3.7c}$$

While (3.7a) is obvious, (3.7b) follows from

$$\sum_{t=p+1}^{n} E(W_{nt}^2 | \mathcal{F}_{t-1}) = \frac{n-p}{n} \mathrm{E}\varepsilon^2_t \frac{1}{n} c^T X^T(n) X(n)c \xrightarrow{a.s.} \eta^2 c^T \Sigma c.$$

Finally, by dominated convergence, $E L_n(\delta) \xrightarrow{n \rightarrow \infty} 0$, which implies (3.7c).

From (3.3) to (3.5) we conclude that

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{d} Z \sim \mathcal{N}(0_{p+1}, \eta^2 \Sigma^{-1}). \tag{3.8}$$

3.2. Simulation study. We present below some numerical results of the proposed model in (3.1) with order $p = 1, 2, 3, 4$ and $T = 30, 80, 100, 500, 1000$ and values of $\alpha_1, \ldots, \alpha_p$ were chosen such that $\sum_{i=1}^{p} \alpha_i < 1/\sqrt{12}$, which ensures finiteness of fourth moments of the count variables; see Lemma 4.1 below.

1000 replications were executed for each of the above combinations. The results consist of the simulated mean estimates of the different parameters, $\theta = (\omega, \alpha_1, \alpha_2, \alpha_3)$, and their corresponding standard errors as deduced from the result (3.8).
The mean estimates of the different parameters for the various time points and orders of the process are consistent to the population values and for the increased number of time points, we could remark a gradual decrease in the standard errors. Since the estimation process is based on the OLS procedure, we could hardly report any non-convergent simulations or computational failures.

| $p$ | $T$ | $\omega = 1.50$ | $\alpha_1 = 0.13$ | $\alpha_2 = 0.08$ | $\alpha_3 = 0.05$ | $\alpha_4 = 0.01$ |
|-----|-----|-----------------|------------------|------------------|------------------|------------------|
| 1   | 30  | 1.251           | 0.098            |                  |                  |                  |
|     |     | (0.321)         | (0.211)          |                  |                  |                  |
|     | 80  | 1.355           | 0.125            |                  |                  |                  |
|     |     | (0.151)         | (0.188)          |                  |                  |                  |
|     | 100 | 1.751           | 0.138            |                  |                  |                  |
|     |     | (0.101)         | (0.124)          |                  |                  |                  |
|     | 500 | 1.442           | 0.134            |                  |                  |                  |
|     |     | (0.087)         | (0.091)          |                  |                  |                  |
|     | 1000| 1.542           | 0.136            |                  |                  |                  |
|     |     | (0.075)         | (0.088)          |                  |                  |                  |
| 2   | 30  | 1.389           | 0.131            | 0.084            |                  |                  |
|     |     | (0.278)         | (0.327)          | (0.209)          |                  |                  |
|     | 80  | 1.477           | 0.134            | 0.077            |                  |                  |
|     |     | (0.150)         | (0.123)          | (0.111)          |                  |                  |
|     | 100 | 1.511           | 0.139            | 0.084            |                  |                  |
|     |     | (0.081)         | (0.099)          | (0.098)          |                  |                  |
|     | 500 | 1.552           | 0.138            | 0.075            |                  |                  |
|     |     | (0.032)         | (0.042)          | (0.038)          |                  |                  |
|     | 1000| 1.467           | 0.139            | 0.082            |                  |                  |
|     |     | (0.022)         | (0.031)          | (0.021)          |                  |                  |
| 3   | 30  | 1.481           | 0.135            | 0.088            | 0.042            |                  |
|     |     | (0.455)         | (0.303)          | (0.276)          | (0.152)          |                  |
|     | 80  | 1.551           | 0.131            | 0.081            | 0.049            |                  |
|     |     | (0.210)         | (0.155)          | (0.101)          | (0.110)          |                  |
|     | 100 | 1.462           | 0.125            | 0.078            | 0.0488           |                  |
|     |     | (0.111)         | (0.101)          | (0.089)          | (0.088)          |                  |
|     | 500 | 1.541           | 0.125            | 0.083            | 0.0521           |                  |
|     |     | (0.088)         | (0.076)          | (0.042)          | (0.034)          |                  |
|     | 1000| 1.4601          | 0.138            | 0.076            | 0.481            |                  |
|     |     | (0.061)         | (0.045)          | (0.034)          | (0.026)          |                  |
| 4   | 30  | 1.495           | 0.128            | 0.075            | 0.053            | 0.009            |
|     |     | (0.323)         | (0.212)          | (0.318)          | (0.176)          | (0.272)          |
|     | 80  | 1.510           | 0.131            | 0.084            | 0.051            | 0.014            |
|     |     | (0.188)         | (0.124)          | (0.232)          | (0.123)          | (0.103)          |
|     | 100 | 1.498           | 0.126            | 0.083            | 0.0487           | 0.009            |
|     |     | (0.092)         | (0.110)          | (0.101)          | (0.075)          | (0.064)          |
|     | 500 | 1.502           | 0.127            | 0.088            | 0.0456           | 0.010            |
|     |     | (0.088)         | (0.064)          | (0.054)          | (0.033)          | (0.042)          |
|     | 1000| 1.489           | 0.126            | 0.083            | 0.490            | 0.009            |
|     |     | (0.052)         | (0.043)          | (0.038)          | (0.018)          | (0.028)          |

Table 1. Simulated Mean estimates and the corresponding standard errors: 100 Simulation experiments
3.3. **A real data example.** As a data example we chose to develop the ongoing expansion of COVID-19 in Mauritius. Mauritius is currently one of the few countries in the world, especially in the African continent, that has achieved a superior recovery rate in the COVID fight in a short lapse of time and has just been referenced in the BBC news recently (see a recent interview in l’Express [l’Express (2020)]). Historically, the first three cases were detected on 18 March 2020 and thereon, the Mauritian government imposes an immediate lockdown on 20 March 2020 followed by a curfew measure on 23.04.2020. In this part of the paper, we analyze the series of infected cases as listed in data.europa.eu. The original data was tested non-stationary by the `adftest` with p-values around 0.0962 and by applying the `diff` routine, we obtain the first-order differenced series that contains a series of positive and negative values while exhibiting severe over-dispersion. Moreover, the `VarSelect` routine yields a superior AIC at lag order 3. Hence, we deem suitable to apply the proposed model in section 3 to the training series from 18.03.2020 to 13.04.2020 and estimate possible forecasts. The ACF and PACF plots of the differenced series are given as:

![Figure 1. Plots of the Infected Covid Cases as at 11.04.2020](image1.png)

![Figure 2. Plots of the Infected Covid Cases as at 11.04.2020](image2.png)

**Remark 3.1 (Heuristics for prediction).** As \( X_t = S_t(\lambda_t) \) with \( \lambda_t \in \mathcal{F}_{t-1} \) (see [3.1]) and \( S_t \) be an i.i.d. sequence of Skellam processes then minimising \( E(E[(X_t - S_t(\lambda))^2|\mathcal{F}_{t-1}]) \) holds for \( \lambda = \lambda_t \). A simple way to fit \( \lambda_t \) is to simply plug in estimates of the parameters and fit it as
follows

\[ \hat{\lambda}_t = \hat{\omega} + \sum_{i=1}^{p} \hat{\alpha}_i X_{t-i}^2. \]

Having in mind the classical mixing techniques (see Doukhan (1994)), this is well known that in case a sample \( X_1, \ldots, X_{t-1} \) is observed then one better uses only a first part of the set of variables \( X_1, \ldots, X_n \) for \( n = n(t) < t \) such that \( n \sim t \) as \( t \to \infty \). In this case \( \hat{\theta} = \hat{\theta}_n \in F_n \) and asymptotic independence yields a convergent procedure. Anyway if we fit now the coefficients over all the training set \( X_1, \ldots, X_{t-1} \) thus the estimates write from (3.2) with sample sizes either \( n \) or \( t-1 \); both estimates may be proved to be close to each other under mild mixing assumptions satisfied here in case e.g. \( t-n = O(\log t) \); for this one makes use of the the classical covariance inequalities Davydov (1970) involving the mixing coefficient \( \beta_X(t-n) \) and \( t \); finally cumbersome calculations yield \( \lim_{t \to \infty} \| \theta(t-1) - \theta(n) \| = 0 \) in probability (in case \( \lim_{t \to \infty} \beta_X(t-n)/t^{2} = 0 \)).

Therefore estimated parameters are \( \hat{\omega} = 0.221(0.108), \hat{\alpha}_1 = 0.142(0.035), \hat{\alpha}_2 = 0.028(0.012), \hat{\alpha}_3 = 0.101(0.042) \). The above results were obtained assuming the training dataset from 18.03.2020 (time is 1) to 19.04.2020 (time \( n \) does not depend on \( t \) here). We now use the above regression estimates or weights to forecast the number of infected COVID cases in Mauritius from 14.04.2020 to 30.04.2020 using the forecasting equation (3.1) in Subsection 3.1, and based on the actual covariate values and compare with the observed cases during the same period.

| Date       | Observed | Forecast | Error |
|------------|----------|----------|-------|
| 20.04.2020 | 3        | 2.812    | 1.812 |
| 21.04.2020 | 0        | 2.011    | 2.011 |
| 22.04.2020 | 0        | 1.321    | 1.321 |
| 23.04.2020 | 1        | 1.342    | 0.342 |
| 24.04.2020 | 2        | 1.818    | 0.182 |
| 25.04.2020 | 0        | 2.031    | 2.031 |
| 26.04.2020 | 0        | 1.302    | 1.302 |
| 27.04.2020 | 1        | 1.312    | 0.312 |
| 28.04.2020 | 0        | 1.404    | 1.404 |
| 29.04.2020 | 0        | 1.323    | 1.323 |
| 30.04.2020 | 0        | 1.235    | 1.235 |

Table 2. supressed Predicted number of Infected Cases in Mauritius, from 20.04.2020 to 30.04.2020

Based on the above results, the Mean Squared Error (MSE) is computed as 1.8534. We unfortunately cannot claim that the current model is enough to describe COVID’s evolution. Since the pandemic seems to be more calm in Mauritius, we think that a zero-inflated model is more adapted: unfortunately in the paper the absolute regularity coefficients are not bounded above, we will report in a further study a more adapted model for COVID in Mauritius. A more global data study is in preparation is a scheme for the progression of the virus over the world. This model is a suggestion among others for modelling the expansion of the disease and some issues of covariates are certainly necessary to include for this difficult question.
4. Proofs and some auxiliary results

4.1. Proofs of the main results.

Proof of Lemma 2.1. Since the probability mass functions of $X$ and $X'$ are symmetric we conclude that these random variables, which are defined by monotone arrangement, have the same sign.

To prove the second statement, suppose w.l.o.g. that $\lambda < \lambda'$. If $Y \sim \text{Skellam}(\lambda)$ and $Z \sim \text{Skellam}(\lambda - \lambda)$ are independent, it follows that $Y + Z \sim \text{Skellam}(\lambda')$. Since the probability mass function of $Y$ is symmetric and unimodal (see e.g. [Alzaid and Omair (2010)]) we have that

$$P(|Y| \leq k) \geq P(|Y + l| \leq k), \quad \forall (l, k) \in \mathbb{Z} \times \mathbb{N}_0,$$

which implies that

$$P(|Y| \leq k) \geq \sum_{l \in \mathbb{Z}} P(|Y + l| \leq k)P(Z = l) = P(|Y + Z| \leq k), \quad \forall k \in \mathbb{N}_0,$$

which means that $|Y|$ is stochastically not greater than $|Y + Z|$. Again by symmetry, we obtain for the above random variables that $|X| \leq |X'|$ holds with probability 1. This implies that $E|X'^2 - X^2| = E[X'^2 - X^2] = 2|\lambda' - \lambda|$.

Proof of Proposition 2.1. (i) We have that

$$\mathbb{E} \Delta_{\gamma,\delta}(Z, Z')$$

$$= \gamma_1 \mathbb{E}|X - X'| + \delta_1 |\lambda - \lambda'| + \sum_{i=2}^{p} \gamma_i |x_i^2 - x_i'^2| + \sum_{j=2}^{q} \delta_j |\lambda_j - \lambda_j'|$$

$$\leq (2\gamma_1 + \delta_1) \left( \sum_{i=1}^{p} \alpha_i y_i + \frac{q}{\alpha_p} \beta_j \right) + \sum_{i=2}^{p} \gamma_i |x_i^2 - x_i'^2| + \sum_{j=2}^{q} \delta_j |\lambda_j - \lambda_j'|.$$

This is guaranteed to be smaller or equal to $\kappa \Delta_{\gamma,\delta}(z, z')$ if

$$(2\gamma_1 + \delta_1) \left( \sum_{i=1}^{p} \alpha_i y_i + \frac{q}{\alpha_p} \beta_j \right) + \sum_{i=2}^{p} \gamma_i y_i - 1 + \sum_{j=2}^{q} \delta_j \lambda_j - 1 \leq \kappa \left( \sum_{i=1}^{p} \gamma_i y_i + \sum_{j=1}^{q} \delta_j \lambda_j \right)$$

is fulfilled for all non-negative $y_1, \ldots, y_p, \lambda_1, \ldots, \lambda_q$. A comparison of coefficients shows that it suffices that the following inequalities are satisfied.

$$(2\gamma_1 + \delta_1) \alpha_1 + \gamma_2 \leq \kappa \gamma_1$$

$$\vdots$$

$$(2\gamma_1 + \delta_1) \alpha_p - 1 + \gamma_p \leq \kappa \gamma_{p-1}$$

$$(2\gamma_1 + \delta_1) \alpha_p \leq \kappa \gamma_p$$

$$(2\gamma_1 + \delta_1) \beta_1 + \delta_2 \leq \kappa \delta_1$$

$$\vdots$$

$$(2\gamma_1 + \delta_1) \beta_{q-1} + \delta_q \leq \kappa \delta_{q-1}$$

$$(2\gamma_1 + \delta_1) \beta_q \leq \kappa \delta_q.$$  \hspace{1cm} (4.1)
We set, w.l.o.g., $2\gamma_1 + \delta_1 = 1$. To proceed, we consider the following system of equations.

\[
\begin{align*}
\alpha_p + \epsilon &= \gamma_p \\
\alpha_{p-1} + \gamma_p + \epsilon &= \gamma_{p-1} \\
&\vdots \\
\alpha_1 + \gamma_2 + \epsilon &= \gamma_1 \\
\beta_q + \epsilon &= \delta_q \\
\beta_{q-1} + \delta_q + \epsilon &= \delta_{q-1} \\
&\vdots \\
\beta_1 + \delta_2 + \epsilon &= \delta_1,
\end{align*}
\]

where $\epsilon = (1 - L)/(2p + q)$. It is obvious that this system of equations has a unique solution with strictly positive $\gamma_1, \ldots, \gamma_p, \delta_1, \ldots, \delta_q$. Adding up the terms on the left-hand and right-hand sides, respectively, we obtain

\[
2 \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j + 2 \sum_{i=2}^{p} \gamma_i + \sum_{j=1}^{q} \delta_j + (2p + q) \epsilon = 2 \sum_{i=1}^{p} \gamma_i + \sum_{j=1}^{q} \delta_j,
\]

which yields that $2\gamma_1 + \delta_1 = 1$, as required. Therefore, we see that with such a choice of $\gamma_1, \ldots, \gamma_p, \delta_1, \ldots, \delta_q$ the following strict inequalities are fulfilled.

\[
\begin{align*}
\alpha_1 + \gamma_2 &< \gamma_1 \\
&\vdots \\
\alpha_{p-1} + \gamma_p &< \gamma_{p-1} \\
\alpha_p &< \gamma_p \\
\beta_1 + \delta_2 &< \delta_1 \\
&\vdots \\
\beta_{q-1} + \delta_q &< \delta_{q-1} \\
\beta_q &< \delta_q.
\end{align*}
\]

Choosing

\[
\kappa = \max \left\{ \frac{\alpha_1 + \gamma_2}{\gamma_1}, \ldots, \frac{\alpha_{p-1} + \gamma_p}{\gamma_{p-1}}, \frac{\alpha_p}{\gamma_p}, \frac{\beta_1 + \delta_2}{\delta_1}, \ldots, \frac{\beta_{q-1} + \delta_q}{\delta_{q-1}}, \frac{\beta_q}{\delta_q} \right\},
\]

we obtain that the system of inequalities (4.1) is satisfied.

(ii) is an immediate consequence of (i).  \qed

**Proof of Proposition 2.2.** Let $Q$ and $Q'$ be arbitrary probability measures supported in $S$ and let $\xi$ be the optimal coupling of $Q$ and $Q'$ w.r.t. the Kantorovich distance, that is,

\[
K_{\gamma, \delta}(Q, Q') = \int_{S \times S} \Delta_{\gamma, \delta}(z, z') \xi(dz, dz').
\]
Then $\xi\pi$ is a coupling of $\pi Q$ and $\pi Q'$ and it follows from Proposition 2.1(i) that
\[
\mathcal{K}(\pi Q, \pi Q') \leq \int \Delta_{\gamma, \delta}(u, u') \xi\pi(du, du')
= \int \left[ \int \Delta_{\gamma, \delta}(u, u') \pi((z, z'), du du') \right] \xi(dz, dz')
\leq \kappa \int \Delta_{\gamma, \delta}(z, z') \xi(dz, dz') = \kappa \mathcal{K}(Q, Q').
\]

**Proof of Theorem 2.1** We consider first the Markov process $(Z_t)_{t \in \mathbb{Z}}$, where
\[
Z_t = (X_t^2, \ldots, X_{t-p+1}^2, \lambda_t, \ldots, \lambda_{t-q+1}).
\]

Let
\[
\mathcal{P} = \left\{ Q : Q \text{ is a probability distribution based in } S, \int_S \sum_{i=1}^{2p} |x_i| Q(dx) < \infty \right\}.
\]

It is well known that the space $\mathcal{P}$ equipped with the Kantorovich metric $\mathcal{K}_{\gamma, \delta}$ is complete. Since by Proposition 2.2 the mapping $\pi$ is contractive it follows by the Banach fixed point theorem that the Markov kernel $\pi$ admits a unique fixed point $Q_0$, i.e. $Q_0 \pi = Q_0$. In other words, $Q_0$ is the unique stationary distribution of the process $(Z_t)_{t \in \mathbb{Z}}$.

Now we consider the process $(Y_t)_{t \in \mathbb{Z}}$, where $Y_t = (X_t, \ldots, X_{t-p+1}, \lambda_t, \ldots, \lambda_{t-q+1})$. Since the conditional distribution of $X_t$ given $X_{t-1}, \lambda_{t-1}, X_{t-2}, \lambda_{t-2}, \ldots$ is, as a Skellam($\lambda_t$) distribution, symmetric about 0, it follows that $Q$ with
\[
Q(\{(-1)^{i_1}k_1 \times \cdots \times (-1)^{i_p}k_p\} \times B_1 \times \cdots \times B_q) = 2^{-p} Q_0(\{k_1^2\} \times \cdots \times \{k_p^2\} \times B_1 \times \cdots \times B_q)
\]
for all $i_1, \ldots, i_p \in \{0, 1\}$, $k_1, \ldots, k_p \in \mathbb{N}_0$ and $B_1, \ldots, B_q \in \mathcal{B}$ is the unique stationary distribution of $(Y_t)_{t \in \mathbb{Z}}$. \hfill \Box

### 4.2. Some auxiliary results.

**Lemma 4.1.** Let $((X_t, \lambda_t))_{t \in \mathbb{Z}}$ be a stationary process satisfying (1.1a) and (3.1), where $\omega, \alpha_1, \ldots, \alpha_p$ are non-negative constants.

(i) If $\alpha = \sum_{i=1}^{n} \alpha_i < \frac{1}{2}$, then $EX_0^2 < \infty$.

(ii) If $\alpha = \sum_{i=1}^{n} \alpha_i < \frac{1}{\sqrt{12}}$, then $EX_0^4 < \infty$.

**Proof of Lemma 4.1** Let $((\tilde{X}_t, \tilde{\lambda}_t))_{t \in \mathbb{N}}$ be Skellam-ARCH process satisfying (1.1a) and (3.1), but with initial values $\tilde{X}_1 = \cdots = \tilde{X}_p = \sqrt{2} \omega$. (The latter condition is imposed to ensure that $E\tilde{X}_1^4, \ldots, E\tilde{X}_p^4$ are guaranteed to be finite.) Since $\tilde{X}_n \xrightarrow{d} X_0$ it follows from Theorem III.6.31 in Pollard [1984, page 58] that we can construct a coupling of these random variables where we have almost sure convergence rather than convergence in probability. Hence, we obtain by Fatou’s lemma that
\[
EX_0^k \leq \liminf_{n \to \infty} E\tilde{X}_n^k, \quad \text{for } k = 2, 4.
\] (4.2)
(i) It follows from (3.1) that
\[ E \bar{X}_t^4 = 2 \tilde{\lambda}_t \leq 2 \left( \omega + \alpha \max\{E \bar{X}_{t-1}^2, \ldots, E \bar{X}_{t-p}^2\} \right). \]

Let \( Z_t = \max\{E \bar{X}_t^2, \ldots, E \bar{X}_{t-p+1}^2\} \). We obtain from the previous display the recursion
\[ Z_t \leq \max\{2(\omega + \alpha Z_{t-1}), Z_{t-1}\}. \]

Therefore,
\[ E \bar{X}_t^2 \leq \frac{2\omega}{1 - 2\alpha}, \]
which yields in conjunction with (3.1) that (i) holds true.

(ii) If \( X \sim \text{Skellam}(\lambda) \), then \( EX^4 = 2\lambda + 12\lambda^2 \). Hence,
\[
E \bar{X}_t^4 = 2 (\omega + \sum_{i=1}^{p} \alpha_i E \bar{X}_{t-i}^2) + 12E \left( \omega + \sum_{i=1}^{p} \alpha_i E \bar{X}_{t-i}^2 \right)^2 \\
\leq 2 \omega + 12 \omega^2 + (2 + 24\omega) \alpha \max\{E \bar{X}_{t-1}^2, \ldots, E \bar{X}_{t-p}^2\} \\
+ 12 \alpha^2 \max\{E \bar{X}_{t-1}^2, \ldots, E \bar{X}_{t-p}^2\}.
\]

With \( \tilde{Z}_t = \max\{E \bar{X}_t^2, \ldots, E \bar{X}_{t-p+1}^2\} \) and \( \bar{\omega} = 2\omega + 12\omega^2 + (2 + 24\omega)\alpha \cdot \frac{2\omega}{1 - 2\alpha} \), we obtain the recursion
\[ \tilde{Z}_t \leq \max\{\bar{\omega} + 12 \alpha^2 \tilde{Z}_{t-1}, \tilde{Z}_{t-1}\}, \]
which leads to
\[ E \bar{X}_t^4 \leq \frac{\bar{\omega}}{1 - 12 \alpha^2}. \]

(ii) follows now from (4.2).

\[ \square \]

Lemma 4.2. Let \(((\bar{X}_t, \tilde{\lambda}_t))_{t \in \mathbb{Z}}\) be a stationary Skellam-ARCH process satisfying (1.1a) and (3.7) and with \( \omega > 0 \). Then the matrix \( \Sigma \) defined in (3.4) is regular.

Proof of Lemma 4.2. We have that
\[ \Sigma = 4E[ZZ^T], \]
where \( Z = (1, X_2, \ldots, X_2^p)^T \).

Assume that \( \Sigma \) is singular: then there exists some \( \gamma = (\gamma_0, \ldots, \gamma_p)^T \neq 0_{p+1} \) such that
\[ 0 = \gamma^T \Sigma \gamma = 4E[(Z^T \gamma)^2], \]
which implies that
\[ P\left(Z^T \gamma = 0\right) = 1. \]

This means that
\[ \gamma_0 + \sum_{i=1}^{p} X_{t-i+1}^2 \gamma_i = 0 \]
holds with probability 1. Since \( \gamma_1 = \cdots = \gamma_p = 0 \) would then imply that \( \gamma = 0_{p+1} \), there exists some \( i_0 \geq 1 \) such that \( \gamma_1 = \cdots = \gamma_{i_0-1} = 0 \) and \( \gamma_{i_0} \neq 0 \). Then
\[ X_{t-i_0}^2 = \frac{1}{\gamma_{i_0}} \left( \gamma_0 + \sum_{i=i_0+1}^{p} \gamma_i X_{t-i+1}^2 \right) \]
that is, $X_{t-i_0}^2$ is fully determined by the past values of the count process. This, however, leads to a contradiction since

$$X_{t-i_0} | \mathcal{F}_{t-i_0-1} \sim \text{Skellam}(\lambda_{t-i_0})$$

with $\lambda_{t-i_0} \geq \omega > 0$. Hence, $\Sigma$ is a regular matrix. □

Acknowledgement We are especially thankful to Bozidar Popovic and Miroslav Ristic for initiating us to investigate Skellam models. Various preliminary discussions and suggestions concerning simple Skellam models were very fruitful and led to other considerations beyond those in the current paper.

This work was funded by CY Initiative of Excellence (grant "Investissements d’Avenir" ANR-16-IDEX-0008), Project "EcoDep", PSI-AAP2020-0000000013.

REFERENCES

ALOMANI, G. A., ALZAID, A. A., AND OMAIR, M. A. (2018). A Skellam GARCH model. *Brazilian Journal of Probability and Statistics* **32**(1), 200–214.

ALZAID, A. A. AND OMAIR, M. A. (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Society*, **33**(1), 17–45.

CHRISTOU, V. AND FOKIANOS, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*. **35**, 55–78.

DAVYDOV, Yu A. (1970). The invariance principle for stationary processes. *Theory of Probability and its Applications*. **15**, 487–498.

DEDECKER, J., DOUKHAN, P., LANG, G., LEÓN, J. R., LOUHICHI S., AND PRIEUR, C. (2007). *Weak Dependence: With Examples and Applications*. Lecture Notes in Statistics **190**, Springer.

DOUC, R., MOULINES, E., PRIOURET, P., AND SOULIER, P. (2018). *Markov Chains*. Springer Series in Operations Research and Financial Engineering.

DOUKHAN, P. (1994). *Mixing, Properties and examples*. Lecture Notes in Statistics **85**, Springer.

DOUKHAN, P., FOKIANOS, K., AND TJIJÆSTHEIM, D. (2012). A new weak dependence condition and applications to moment inequalities. *Statistic and Probability Letters* **82**(5), 942–948.

DOUKHAN, P., LEUCHT, A., AND NEUMANN, M. H. (2020). Mixing properties of non-stationary INGARCH(1,1) processes. Manuscript.

DOUKHAN, P. AND LOUHICHI, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications* **84**(2), 313–342.

DOUKHAN, P. AND NEUMANN, M. H. (2008). The notion of $\psi$-weak dependence and its applications to bootstrapping time series. *Probability Surveys* **5**, 146–168.

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

DOUKHAN, P. AND WINTENBERGER, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications* **118**, 1997–2013.

EBERLE, A. (2019). Markov processes, Lecture Notes University of Bonn. [https://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Andreas_Eberle/MarkovProcesses1920/MarkovProcesses1920.pdf](https://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Andreas_Eberle/MarkovProcesses1920/MarkovProcesses1920.pdf)

HALL, P. AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press.
Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution. *Journal of the Royal Statistical Society, Series A*. **100**, 415-416.

Jugnauth, P. (2020). The preparation of a Covid-19 Bill evoked after several safe days. *L’Express, SFR Presse. Paris*. April 18, [https://www.lexpress.mu/article/375179/pravind-jugnauth-interviewe-bbc-preparation-dun-covid-19-bill-evoquee](https://www.lexpress.mu/article/375179/pravind-jugnauth-interviewe-bbc-preparation-dun-covid-19-bill-evoquee)

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

Pollard, D. (1980). *Convergence of Stochastic Processes*. Springer, New York.

Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society, Series A*. **109**(3), 296.