International Journal of Forecasting 40 (2024) 1359–1390

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Generalized Poisson difference autoregressive processes

Giulia Carallo [a,*], Roberto Casarin [a], Christian P. Robert [b,c]

[a] Ca' Foscari University of Venice, Italy
[b] Université Paris-Dauphine, France
[c] University of Warwick, United Kindom

ARTICLE INFO

Keywords:
Bayesian inference
Counts time series
Cyber risk
GARCH models
Poisson processes

ABSTRACT

This paper introduces a novel stochastic process with signed integer values. Its autoregressive dynamics effectively captures persistence in conditional moments, rendering it a valuable feature for forecasting applications. The increments follow a Generalized Poisson distribution, capable of accommodating over- and under-dispersion in the conditional distribution, thereby extending standard Poisson difference models. We derive key properties of the process, including stationarity conditions, the stationary distribution, and conditional and unconditional moments, which prove essential for accurate forecasting. We provide a Bayesian inference framework with an efficient posterior approximation based on Markov Chain Monte Carlo. This approach seamlessly incorporates inherent parameter uncertainty into predictive distributions. The effectiveness of the proposed model is demonstrated through applications to benchmark datasets on car accidents and an original dataset on cyber threats, highlighting its superior fitting and forecasting capabilities compared to standard Poisson models.

1. Introduction

Integer-valued variables are ubiquitous in diverse fields, including medicine (Cardinal, Roy, & Lambert, 1999), epidemiology (Davis, Dunsmuir, & Wang, 1999; Zeger, 1988), finance (Liesenfeld, Nolte, & Pohlmeier, 2006; Rydberg & Shephard, 2003), and economics (Freeland, 1998; Freeland & McCabe, 2004). While literature is abundant on modelling integer-valued data, fewer studies specifically address signed integer-valued data, as their modelling presents unique challenges (e.g., see Cunha, Vasconcellos, & Bourguignon, 2018; Koopman, Lit, & Lucas, 2017; Pedeli & Karlis, 2011; Shahtahmassebi & Moyeed, 2014, 2016). This paper contributes to this literature by introducing a novel dynamic model based on the generalized Poisson difference (GPD) distribution. The GPD family, pioneered by Consul (1986), is highly flexible, accommodating unevenly dispersed data, both over- and under-dispersed. It encompasses equal dispersion as a special case, recovering the standard Poisson difference (PD) distribution, also known as the Skellam distribution (see, e.g., Shahtahmassebi and Moyeed (2016)).

In the literature on integer-valued autoregressive processes, two primary modelling approaches have emerged: integer-valued autoregressive-moving average models (INARMA) and integer-valued GARCH (INGARCH). INARMA models, introduced through binomial thinning (Al-Osh & Alzaid, 1987; McKenzie, 1986), have been extended in various directions (see, e.g., Alzaid & Al-Osh, 1993; Jin-Guan & Yuan, 1991), including processes defined on signed integers (e.g., Alzaid & Omair, 2014; Andersson & Karlis, 2014; Cunha et al., 2018; Freeland, 2010; Kim & Park, 2008). The INGARCH process with Poisson margins was initially proposed by Ferland, Latour, and Oraichi (2006) and subsequently extended by Zhu (2012) to accommodate under- and overdispersion. Additionally, Koopman et al. (2017) and Alomani, Alzaid, Omair, et al. (2018) extended it to handle signed integers.

* Corresponding author.
E-mail addresses: giulia.carallo@unive.it (G. Carallo), r.casarin@unive.it (R. Casarin), Christian.Robert@ceremade.dauphine.fr (C.P. Robert).

https://doi.org/10.1016/j.ijforecast.2023.11.009

This paper adopts the INGARCH modelling approach due to its numerous advantages. Firstly, it demonstrates improved fitting for count time series in the presence of over- and under-dispersion and time-varying variance (Zhu, 2012). Secondly, it exhibits simpler conditional probabilities, rendering it more tractable for likelihood-based inference methods and forecasting. We propose a new INGARCH model with Generalized Poisson Difference (GPD) conditional probabilities, referred to as GPD-INGARCH. This model extends PD-INGARCH models (Koopman et al., 2017) to account for uneven dispersion and GP-INGARCH models (Zhu, 2012) to accommodate signed integer data. Additionally, we investigate the theoretical properties of the proposed GPD-INGARCH and develop a suitable inference procedure.

Thinning operators play a crucial role in analyzing the properties of integer-valued processes. Among these operators, binomial thinning, initially introduced by Steutel and van Harn (1979), is the most commonly used. Over time, it has been subject to generalizations in various directions (see, e.g., Alzaid & Al-Osh, 1993; Kim & Park, 2008; Latour, 1998; McKenzie, 1985, 1986; Osh & Aly, 1992). Detailed reviews of binomial thinning and its generalizations can be found in works by Scotto, Weiß, and Gouveia (2015) and Weiß (2008). Thinning operations can be combined linearly to define new operations such as the binomial thinning difference (Freeland, 2010) and the quasi-binomial thinning difference (Cunha et al., 2018). This paper uses the quasi-binomial thinning difference to establish essential process properties, including stationarity conditions and moments.

Another significant contribution pertains to the inference approach. While maximum likelihood estimation has been extensively explored for integer-valued processes, Bayesian inference procedures have received comparatively less attention. Notably, Chen and Lee (2016) introduced the Bayesian zero-inflated GP-INGARCH model with structural breaks, Zhu and Li (2009) proposed a Bayesian Poisson INGARCH(1,1), and Chen, So, Li, and Sriboonchitta (2016) presented a Bayesian Autoregressive Conditional Negative Binomial model. In this study, we develop a Bayesian inference procedure for the proposed GPD-INGARCH process, employing a Markov Chain Monte Carlo (MCMC) algorithm for posterior approximation. One of the advantages of the Bayesian approach is that extra-sample information and constraints on the parameter value can be easily included in the estimation process through the prior distributions. For non-Gaussian GARCH models, Bayesian inference can be effectively combined with data augmentation strategies to enhance the tractability of the likelihood function and numerical methods for addressing complex inference problems (see Ardia (2008)). Furthermore, the Bayesian framework allows for the straightforward incorporation of parameter uncertainty into the predictive distribution, thereby enhancing the robustness of forecasts (see, e.g., McCabe & Martin, 2005; McCabe, Martin, & Harris, 2011, and references therein).

We apply our model to a benchmark dataset on car accidents near the Schiphol airport, previously studied in Andersson and Karlis (2014), and Brijs, Karlis, and Wets (2008) with the primary purpose of evaluating fitting and forecasting abilities of our model with respect to models previously used for this dataset. We also cover a second application to a cyber-threat dataset. Cyber threats are increasingly considered a top global risk for the financial and insurance sectors and the economy as a whole (e.g. EIOPA, 2019). As highlighted by Hassanien et al. (2016), the frequency of cyber events has markedly increased in recent years, with cyber-attacks occurring daily. Understanding the dynamics of cyber threats and their impact is crucial for ensuring effective controls and risk mitigation tools. Despite the significance of this issue, research on the analysis of cyber threats is limited and dispersed across various fields, including cyber security (Agrafiotis, Nurse, Goldsmith, Creese, & Upton, 2018), criminology (Brenner, 2004), economics (Anderson & Moore, 2006), and sociology. Only a few studies have delved into statistical modelling and forecasting of cyber-attacks. Xu, Hua, and Xu (2017) introduced a copula model to predict cyber-security effectiveness, while (Werner, Yang, & McConky, 2017) utilized an autoregressive integrated moving average model to forecast the daily number of cyber-attacks. Moreover, Edwards, Hofmeyr, and Forrest (2015) applied Bayesian Poisson and negative binomial models to analyze data breaches, revealing evidence of over-dispersion and the absence of time trends in the number of breaches. For a comprehensive review of modelling cyber threats, refer to Husák, Komárková, Bou-Harb, and Čeleda (2018). We bring a new perspective by providing evidence of temporal patterns in the mean and variance of the threats, which can be used to predict threat arrivals. In-sample and out-of-sample forecasting comparisons between alternative model specifications are carried out to provide insights into the threats' dynamics and over-dispersion features.

The paper is organized as follows. Section 2 introduces both the GPD distribution and the GPD-INGARCH process. Section 3 presents some process properties. Section 4 provides a full Bayesian inference procedure. Sections 5 and 6 illustrate GPD-INGARCH on simulated and real-world data. Section 7 concludes.

2. Generalized Poisson difference INGARCH

A random variable X follows a Generalized Poisson (GP) distribution if its probability mass function (pmf) is

$$P_x(\theta, \lambda) = \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-\theta - x\lambda}, \qquad x = 0, 1, 2, \ldots \quad (1)$$

with rate parameter $\theta > 0$ and skewness parameter $0 \leq \lambda < 1$ (Consul, 1986). We denote this distribution as $GP(\theta, \lambda)$. Let $X \sim GP(\theta_1, \lambda)$ and $Y \sim GP(\theta_2, \lambda)$ be two independent GP random variables where $\theta_1 = (\sigma^2 + \mu)/2$ and $\theta_2 = (\sigma^2 - \mu)/2$, with $\sigma^2 > 0$ and $\mu \in \mathbb{R}$ such that $\theta_1 > 0$ and $\theta_2 > 0$. As proved in Consul (1986), the probability distribution of $Z = (X - Y)$ follows a Generalized Poisson Difference (GPD) distribution with

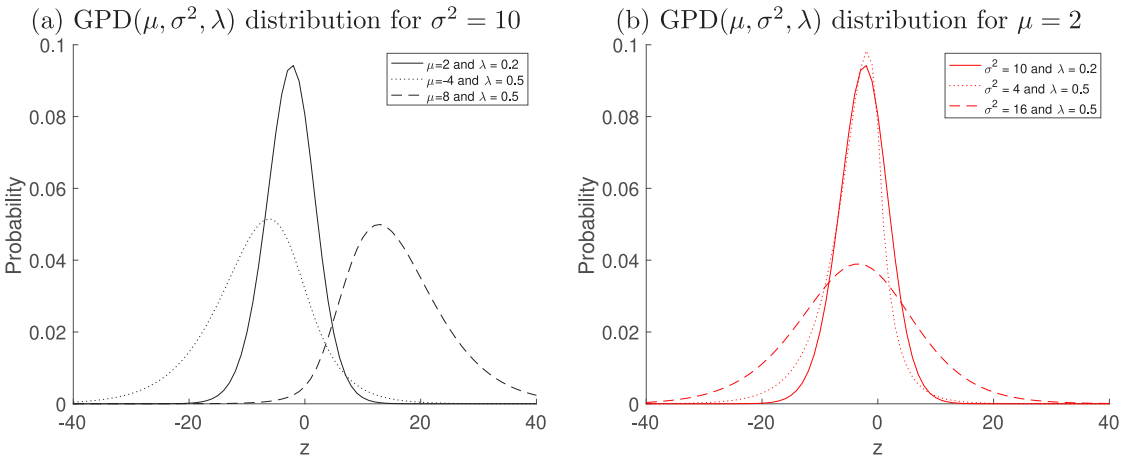

Fig. 1. Generalized Poisson difference GPD(μ, σ^2, λ) pmf for some values of λ, μ and σ^2. The pmf with $\lambda = 0.2$, $\mu = 2$, and $\sigma^2 = 10$ (solid line) is taken as a baseline in both panels.

pmf:

$$P_z(\mu, \sigma^2, \lambda) = e^{-\sigma^2 - z\lambda} \sum_{s=\max(0,-z)}^{+\infty} \frac{1}{4} \frac{\sigma^4 + \mu^2}{s!(s+z)!} \left[\frac{\sigma^2 + \mu}{2} + (s+z)\lambda\right]^{s+z-1} \left[\frac{\sigma^2 - \mu}{2} + s\lambda\right]^{s-1} e^{-2\lambda s}, \quad (2)$$

where σ^2, μ, and λ are the scale, location, and skewness parameters, respectively. We denote this distribution with $GPD(\mu, \sigma^2, \lambda)$.

The moments of a GPD random variable can be derived in closed form by leveraging the representation of the GPD as the difference between independent GP random variables.

Lemma 1. *Let $Z \sim GPD(\mu, \sigma^2, \lambda)$, then its mean and variance are given by:*

$$\mathbb{E}(Z) = \frac{\mu}{1-\lambda}, \qquad \mathbb{V}(Z) = \frac{\sigma^2}{(1-\lambda)^3}. \quad (3)$$

The Pearson skewness and kurtosis are also expressed as follows:

$$S(Z) = \frac{\mu}{\sigma^3} \frac{(1+2\lambda)}{\sqrt{1-\lambda}}, \qquad K(Z) = 3 + \frac{1 + 8\lambda + 6\lambda^2}{\sigma^2(1-\lambda)}. \quad (4)$$

See Appendix A for a proof.

Fig. 1 illustrates the sensitivity of the probability distribution to the location μ (panel a) and scale σ^2 (panel b) parameters. The various lines in each plot demonstrate the impact of the skewness parameter λ. For given values of λ and μ, when σ^2 decreases, the dispersion of the GPD decreases (dotted and dashed lines, right plot). For given values of λ and σ^2, the distribution is right-skewed for $\mu = 8$, which corresponds to $S(Z) = 0.7155$, and left-skewed for $\mu = -4$, which corresponds to $S(Z) = -0.3578$, (dotted and dashed lines, left plot). Further numerical illustrations can be found in the Supplementary Material.

In contrast with the usual GARCH(p, q) process (Francq & Zakoian, 2019), the INGARCH(p, q) process is defined as an integer-valued process $\{Z_t\}_{t\in\mathbb{Z}}$, where Z_t is a series of counts. Let $\mathcal{F}_{t-1}$ be the σ-field generated by $\{Z_{t-j}\}_{j\geq 1}$, then the GPD-INGARCH(p, q) is defined as

$$Z_t|\mathcal{F}_{t-1} \sim GPD(\mu_t, \sigma_t^2, \lambda)$$

with

$$\mu_t = (1-\lambda)\alpha_0 + (1-\lambda)\sum_{i=1}^{p} \alpha_i Z_{t-i} + \sum_{j=1}^{q} \beta_j \mu_{t-j}, \quad (5)$$

where $\alpha_0 \in \mathbb{R}$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $i = 1, \ldots, p$, $p \geq 1$, $j = 1, \ldots, q$, $q \geq 0$. For $q = 0$, the model reduces to a GPD-INARCH(p) model and for $\lambda = 0$, one obtains a Skellam INGARCH(p, q) model which extends to Poisson differences, the Poisson INGARCH(p, q) model of Ferland et al. (2006). Based on (3), the conditional mean and variance of the process are:

$$\mathbb{E}(Z_t|\mathcal{F}_{t-1}) = \frac{\mu_t}{1-\lambda}, \quad \mathbb{V}(Z_t|\mathcal{F}_{t-1}) = \frac{\sigma_t^2}{(1-\lambda)^3}, \quad (6)$$

respectively. Simulated GPD-INGARCH sequences can be obtained as differences between GP sequences $Z_t = X_t - Y_t$ with $X_t \sim GP(\theta_{1t}, \lambda)$, $Y_t \sim GP(\theta_{2t}, \lambda)$ where $\theta_{1t} = (\sigma_t^2 + \mu_t)/2$ and $\theta_{2t} = (\sigma_t^2 - \mu_t)/2$. Each random sequence is generated by the branching method of Famoye (1997), which performs faster than the inversion method for large values of θ_{1t} and θ_{2t}. To measure overdispersion we assume $\sigma_t^2 = |\mu_t|\phi(1-\lambda)^2$ where ϕ is an overdispersion parameter. There is overdispersion in the conditional distribution, that is $E(Z_t|\mathcal{F}_{t-1})/V(Z_t|\mathcal{F}_{t-1}) < 1$, if $\phi > 1$. This condition is always satisfied for $0 \leq \lambda < 1$ since $\phi > (1-\lambda)^{-2}$ in order to have $\theta_{2t} > 0$ and a well defined GPD distribution. When $\lambda < 0$, the GPD is still well defined provided $\lambda > \max\{\max(-1, -\theta_j/m_j), j = 1, 2\}$ (see Supplementary Material) and both under and overdispersion are allowed in our GPD-INGARCH model.

Fig. 2 provides some simulated examples of the GPD-INGARCH(1, 1) process for different values of α_0, α_1 and β_1. We consider two parameter settings: low persistence, that is $\alpha_1 + \beta_1$ much less than 1 (first column in Fig. 2), and high persistence, that is $\alpha_1 + \beta_1$ close to 1 (second

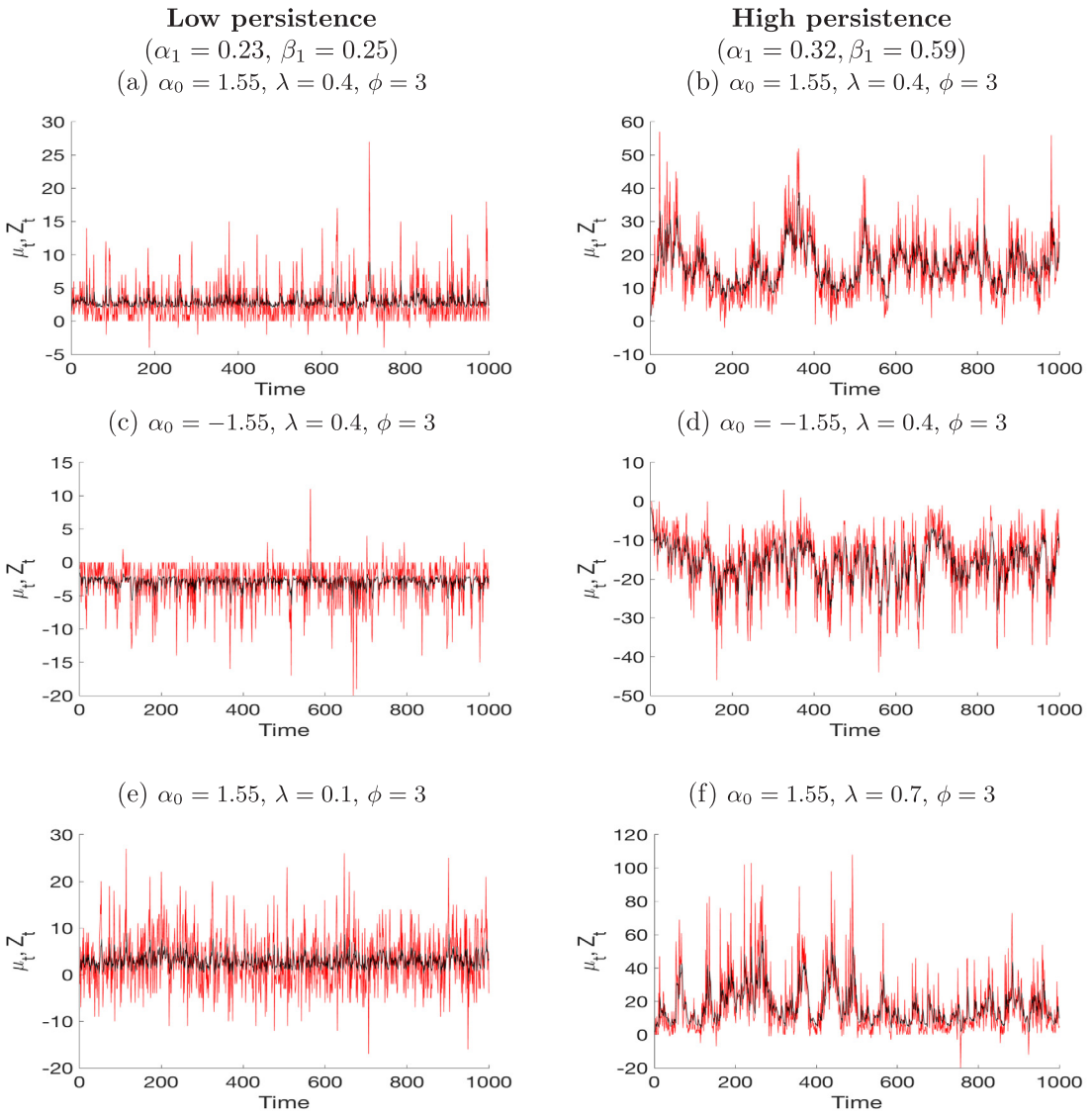

Fig. 2. Simulated INGARCH(1, 1) paths for different values of the parameters α_0, α_1 and β_1. In Panels from (a) to (d), the effect of α_0 ($\alpha_0 > 0$ in the first line and $\alpha_0 < 0$ in the second line) is illustrated with $\lambda = 0.4$ and $\phi = 3$. In Panels (e) and (f), the effect of λ ($\lambda = 0.1$ left and $\lambda = 0.7$ right) is produced in both settings.

column in Fig. 2). The first and second rows show paths for a positive and negative value of the intercept α_0, respectively. The last row illustrates the effect of λ on the trajectories with respect to the baselines in panels (a) and (b). By comparing (I.a) and (I.b) in Fig. 3, one can observe that an increase in β_1 leads to higher serial correlation and the kurtosis levels (compare (II.a) and (II.b)).

Stationarity is a crucial property for forecasting applications. Therefore, we present a necessary condition on the parameters α_i and β_j to ensure that a second-order stationary process has an INGARCH representation. Define the two following polynomials: $D(B) = 1-\beta_1 B-\cdots-\beta_q B^q$ and $G(B) = \alpha_1 B + \cdots + \alpha_p B^p$, where B is the backshift operator. Assume the $D(z)$ roots lie outside the unit circle. For non-negative β_j this is equivalent to assume $D(1) = \sum_{j=1}^{q} \beta_j < 1$. Consequently, the operator $D(B)$ has inverse $D^{-1}(B)$ and it is possible to write $\mu_t = D^{-1}(B)(\alpha_0 + G(B)Z_t) = \alpha_0 D^{-1}(1) + H(B)Z_t$, where $H(B) = G(B)D^{-1}(B) = \sum_{j=1}^{\infty} \psi_j B^j$ and the ψ_j are given by the power expansion of the rational function $G(z)/D(z)$ in a neighbourhood of zero. Let $K(B) = D(B) - G(B)$, then the necessary condition can be written as follows:

Proposition 1. *A necessary condition for a second-order stationary process $\{Z_t\}_{t\in\mathbb{Z}}$ to satisfy Eq. (5) is that $K(1) = D(1) - G(1) > 0$ or equivalently $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$.*

Proof. See Appendix A □

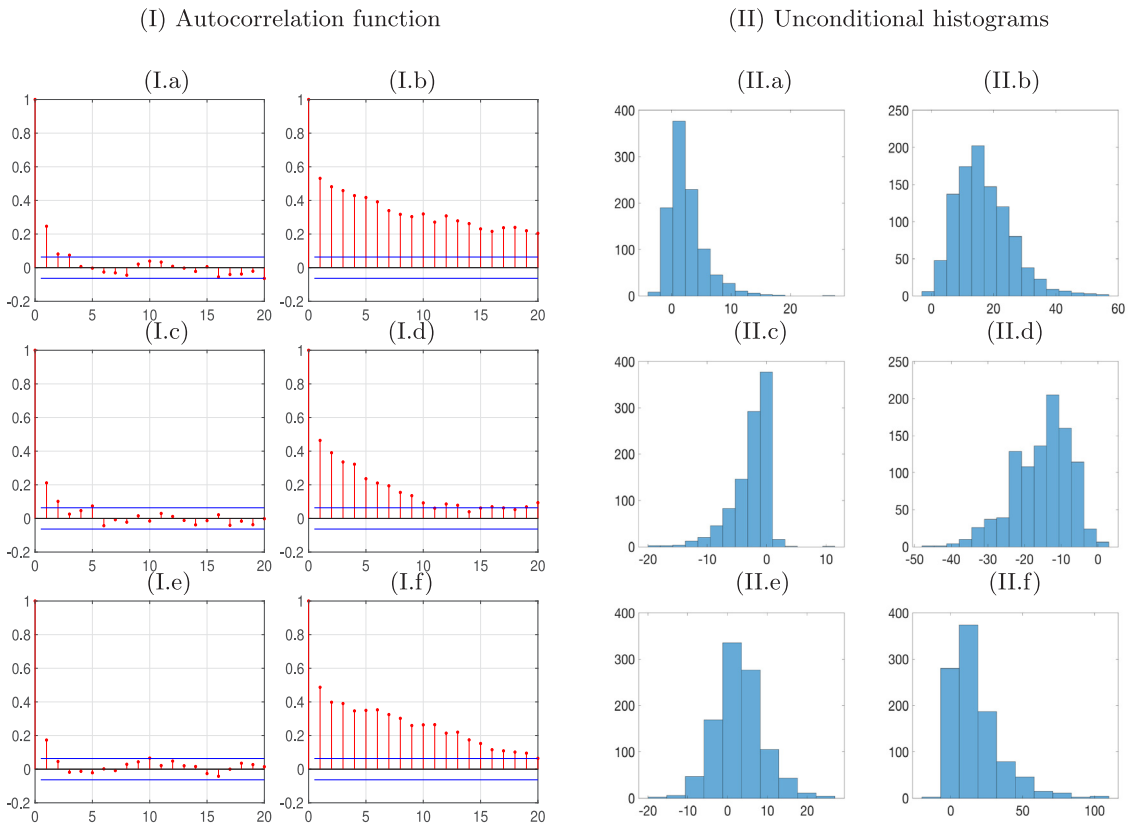

Fig. 3. Autocorrelation functions (Panel I) and unconditional distributions (Panel II) of $\{Z_t\}_{t\in\mathcal{Z}}$ for the different cases presented in Fig. 2 (different columns in each panel).

3. Properties of the GPD-INGARCH

We study here some relevant properties of our process by providing a suitable thinning representation and following the strategy used in Ferland et al. (2006) and Zhu (2012) for Poisson INGARCH and Generalized Poisson INGARCH, respectively. We use the quasi-binomial thinning as defined in Weiß (2008) and the thinning difference (Cunha et al., 2018) operators. Some background material and preliminary results, together with the proofs of the results of this section, are given in Appendix A.

3.1. Thinning representation and stationarity

We now show how the INGARCH process can be obtained as a limit of successive approximations. Let $\{U_{1t}\}_{t\in\mathbb{Z}}$ and $\{U_{2t}\}_{t\in\mathbb{Z}}$ be two independent sequences of independent GP random variables. For each $t\in\mathbb{Z}$ and $i\in\mathbb{N}$, let $\{V_{1t,i,j}\}_{j\in\mathbb{N}}$ and $\{V_{2t,i,j}\}_{j\in\mathbb{N}}$ be two independent sequences of independent integer random variables. Moreover, assume that U_s and $V_{t,i,j}$, $s,t\in\mathbb{Z}$, $i,j\in\mathbb{N}$, are mutually independent variables, and define the sequence:

$$X_t^{(n)} = \mathbb{I}(n=0)(1-\lambda)U_{1t} + (1-\lambda)U_{1t} + \mathbb{I}(n>0)(1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{A_{i,t}^n} V_{1t-i,i,j} \tag{7}$$

$$Y_t^{(n)} = \mathbb{I}(n=0)(1-\lambda)U_{2t} + \mathbb{I}(n>0)(1-\lambda)U_{2t} + (1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{B_{i,t}^n} V_{2t-i,i,j}, \tag{8}$$

where $A_{i,t}^n = X_{t-i}^{(n-i)}/(1-\lambda)$, and $B_{i,t}^n = (Y_{t-i}^{(n-i)})/(1-\lambda)$.

In the following, we introduce a new thinning difference operator and show that $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$ has this thinning representation. See Cunha et al. (2018) for an application of the thinning operation to GPD-INAR processes and Supplementary Material for further details.

Definition 1. Let $X\sim GP(\theta_1,\lambda)$ and $Y\sim GP(\theta_2,\lambda)$ be two independent GP random variables, and $Z=X-Y$, with $\mu=\theta_1-\theta_2$ and $\sigma^2=\theta_1+\theta_2$. We define the new operator $\diamond$as:

$$\rho\diamond Z|Z \overset{\mathrm{d}}{=} (\rho_{\theta_1,\lambda}\circ X) - (\rho_{\theta_2,\lambda}\circ Y)|(X-Y), \tag{9}$$

where $(\rho_{\theta_1,\lambda}\circ X)$ and $(\rho_{\theta_2,\lambda}\circ Y)$ are the quasi-binomial thinning operations such that $(\rho_{\theta_1,\lambda}\circ X)|X=x\sim QB(p,\lambda/\theta_1,x)$ and $(\rho_{\theta_2,\lambda}\circ Y)|Y=y\sim QB(p,\lambda/\theta_2,y)$. The symbol "$A\overset{\mathrm{d}}{=}B$" means that both random variables A and B have the same distribution.

Using this new operator, we can represent $Z_t^{(n)}$ as follows.

Proposition 2. *The process* $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$ *has the representation:*

$$Z_t^{(n)} = (1-\lambda)U_t + (1-\lambda)^2 \sum_{i=1}^{n} \varphi_i^{(t-i)} \diamond \left(\frac{Z_{t-i}^{(n-i)}}{1-\lambda} \right), \quad n > 0, \tag{10}$$

$Z_t^{(n)} = (1-\lambda)U_t$ *for* $n = 0$ *and* $Z_t^{(n)} = 0$ *for* $n < 0$, *where* $\varphi_i^{(\tau)} \diamond$ *indicates the sequence of random variables with mean* $\psi_i/(1-\lambda)$, *involved in the thinning operator at time* τ *and* $\{U_t\}_{t\in\mathbb{Z}}$ *is a sequence of independent GPD random variables with mean* $\psi_0/(1-\lambda)$ *with* $\psi_0 = \alpha_0/D(1)$.

The proposition above shows that $\{Z_t^{(n)}\}_{t\in\mathbb{Z}}$ is obtained through a cascade of thinning operations along the sequence $\{U_t\}_{t\in\mathbb{Z}}$ and is a finite weighted sum of independent GPD random variables. It follows that both the expected value and the variance of $Z_t^{(n)}$ are well defined. Moreover, it can be seen that $\mathbb{E}[Z_t^{(n)}]$ does not depend on t; hence it can be denoted with μ_n. Using Proposition 2 and $\mu_k = 0$ if $k < 0$, it is possible to write μ_n as follows

$$\begin{aligned}\mu_n &= (1-\lambda)\mathbb{E}(U_t) + (1-\lambda)^2 \sum_{i=1}^{n} \mathbb{E}\left(\varphi_i^{(t-i)} \diamond \left(\frac{Z_{t-i}^{(n-i)}}{1-\lambda} \right) \right) \\ &= \psi_0 + \sum_{j=1}^{\infty} \psi_j \mu_{n-j} = D^{-1}(B)\alpha_0 + H(B)\mu_n,\end{aligned} \tag{11}$$

from which it follows that $D(B)\mu_n = G(B)\mu_n + \alpha_0 \Leftrightarrow K(B)\mu_n = \alpha_0$, where $K(B) = D(B) - G(B)$. The last equation shows that the sequence $\{\mu_n\}_{n\in\mathbb{N}}$ satisfies a finite difference equation with constant coefficients. The characteristic polynomial is $K(z)$, and all its roots lie outside the unit circle if $K(1) > 0$. Under this assumption, one can prove the following.

Proposition 3. *If* $K(1) > 0$ *then:*

(i) $\{Z_t^{(n)}\}_{n\in\mathbb{N}}$ *has an almost sure limit;*
(ii) $\{Z_t^{(n)}\}_{n\in\mathbb{N}}$ *has a mean-square limit;*
(iii) $\{Z_t^{(n)}\}_{t\in\mathbb{Z}}$ *is strictly stationary, for any given n.*

Since $\{Z_t\}_{t\in\mathbb{Z}}$ is the almost sure limit of $\{Z_t^{(n)}\}_{t\in\mathbb{Z}}$ one gets the following result:

Proposition 4. *The process* $\{Z_t\}_{t\in\mathbb{Z}}$ *is strictly stationary and admits finite first- and second-order moments.*

Furthermore, one can find the conditional distribution of $\{Z_t\}_{t\in\mathbb{Z}}$ using the properties of the thinning representation of Proposition 2.

Proposition 5. *Let* $\mathcal{F}_{t-1} = \sigma(\{Z_u\}_{u\le t-1})$, *for* $t \in \mathbb{Z}$, *the conditional law of* $\{Z_t^{(n)}\}_{t\in\mathbb{Z}}$ *given* $\mathcal{F}_{t-1}$ *converges to a* $GPD(\mu_t, \sigma_t^2, \lambda)$.

3.2. Moments of the GPD-INGARCH

The conditional mean and variance of the process Z_t can be easily derived from Eq. (6), while the unconditional mean and variance are

$$\mathbb{E}(Z_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j}, \quad \mathbb{V}(Z_t) = \phi^3\mathbb{E}(\sigma_t^2) + \mathbb{V}(\mu_t), \tag{12}$$

where $\phi = 1/(1-\lambda)$. A set of equations exists from which the variance and autocorrelation function of the process can be obtained. Suppose Z_t follows the INGARCH(p,q) model in Eq. (5) with $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 0$. From Th. 1 part (iii) in Weiß (2009), the autocovariance $\gamma_Z(k) = \mathbb{C}ov[Z_t, Z_{t-k}]$ and $\gamma_\mu(k) = \mathbb{C}ov[\mu_t, \mu_{t-k}]$ satisfy the linear equations

$$\begin{aligned}\gamma_Z(k) &= \sum_{i=1}^{p}\alpha_i\gamma_Z(|k-i|) + \sum_{j=1}^{\min(k-1,q)}\beta_j\gamma_Z(k-j) \\ &\quad + \sum_{j=k}^{q}\beta_j\gamma_\mu(j-k), \quad k \ge 1;\end{aligned} \tag{13}$$

$$\begin{aligned}\gamma_\mu(k) &= \sum_{i=1}^{\min(k,p)}\alpha_i\gamma_\mu(|k-i|) + \sum_{i=k+1}^{p}\alpha_i\gamma_Z(i-k) \\ &\quad + \sum_{j=1}^{q}\beta_j\gamma_\mu(|k-j|), \quad k \ge 0.\end{aligned} \tag{14}$$

An explicit expression of these moments is available for two special cases.

Example 1 (*GPD-INARCH(1)*)**.** Consider the GPD-INARCH(1) model with $\mu_t = \alpha_0 + \alpha_1 Z_{t-1}$, then the linear equations in Eq. (14), become

$$\begin{aligned}\gamma_Z(k) &= \sum_{i=1}^{p}\alpha_i\gamma_Z(|k-i|) + \delta_{k0}\mu, \quad k \ge 0; \\ \gamma_\mu(k) &= \sum_{i=1}^{\min(k,p)}\alpha_i\gamma_\mu(|k-i|) + \sum_{i=k+1}^{p}\alpha_i\gamma_Z(i-k), \quad k \ge 0,\end{aligned}$$

where the second equation comes from Example 2 in Weiß (2009). We derive the following autocovariances

$$\gamma_Z(k) = \begin{cases}\alpha_1^{k-1}\gamma_Z(1), & \text{for } k \ge 2 \\ \alpha_1(\phi^3\mathbb{E}(\sigma_t^2)) + \alpha_1\mathbb{V}(\mu_t), & \text{for } k = 1,\end{cases} \tag{15}$$

$$\gamma_\mu(k) = \begin{cases}\alpha_1^k\mathbb{V}(\mu_t), & \text{for } k \ge 1 \\ \alpha_1^2(\phi^3\mathbb{E}(\sigma_t^2)) + \alpha_1^2\mathbb{V}(\mu_t), & \text{for } k = 0\,.\end{cases} \tag{16}$$

The variance of μ_t and of Z_t and the autocorrelation functions are

$$\begin{aligned}\mathbb{V}(\mu_t) &= \frac{\alpha_1^2(\phi^3\mathbb{E}(\sigma_t^2))}{1-\alpha_1^2}, \quad \mathbb{V}(Z_t) = \frac{\phi^3\mathbb{E}(\sigma_t^2)}{1-\alpha_1^2}, \\ \rho_\mu(k) &= \alpha_1^k, \quad \rho_Z(k) = \alpha_1^k,\end{aligned} \tag{17}$$

respectively, where $\phi = 1/(1-\lambda)$.

Example 2 (*GPD-INGARCH(1,1)*)**.** Consider the GPD-INGARCH(1,1) model with $\mu_t = \alpha_0 + \alpha_1 Z_{t-1} + \beta_1\mu_{t-1}$.

From Eq. (14), it follows

$$\gamma_Z(k) = \begin{cases} (\alpha_1+\beta_1)^{k-1}\gamma_Z(1), & \text{for } k \geq 2 \\ \alpha_1(\phi^3\mathbb{E}(\sigma_t^2)) + (\alpha_1+\beta_1)\mathbb{V}(\mu_t), & \text{for } k = 1\,. \end{cases} \tag{18}$$

We can now determine $\mathbb{V}(\mu_t)$. First, note that we have

$$\gamma_\mu(k) = \begin{cases} (\alpha_1+\beta_1)^{k}\mathbb{V}(\mu_t), & \text{for } k \geq 1 \\ \alpha_1^2[\phi^3\mathbb{E}(\sigma_t^2)] + (\alpha_1+\beta_1)^2\mathbb{V}(\mu_t), & \text{for } k = 0, \end{cases} \tag{19}$$

where the second line in Eq. (19) is equal to $\mathbb{V}(\mu_t)$. From this latter equation and Eq. (12), we can derive the following expressions

$$\begin{aligned} \mathbb{V}(\mu_t) &= \frac{\alpha_1^2(\phi^3\mathbb{E}(\sigma_t^2))}{1-(\alpha_1+\beta_1)^2}, \\ \mathbb{V}(Z_t) &= \frac{\phi^3\mathbb{E}(\sigma_t^2)(1-(\alpha_1+\beta_1)^2+\alpha_1^2)}{1-(\alpha_1+\beta_1)^2}. \end{aligned} \tag{20}$$

The autocorrelations functions are

$$\begin{aligned} \rho_\mu(k) &= (\alpha_1+\beta_1)^k, \\ \rho_Z(k) &= (\alpha_1+\beta_1)^{k-1}\,\frac{\alpha_1[1-\beta_1(\alpha_1+\beta_1)]}{1-(\alpha_1+\beta_1)^2+\alpha_1^2}. \end{aligned} \tag{21}$$

4. Bayesian inference

We propose a Bayesian framework to estimate GPD-INGARCH models that exploits the stochastic representation of the GPD as a difference between latent GP variables and the data augmentation principle to make both the likelihood function and the posterior distribution more tractable. The Bayesian approach to prediction includes parameter uncertainty in the posterior predictive, thus providing robust forecasts.

4.1. Prior assumption

We assume the following prior distributions. A Dirichlet prior distribution is chosen for $\boldsymbol{\varphi} = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$, i.e., $\boldsymbol{\varphi} \sim Dir_{d+1}(c)$, with density:

$$\pi(\boldsymbol{\varphi}) = \frac{\Gamma\left(\sum_{i=0}^d c_i\right)}{\prod_{i=0}^d \Gamma(c_i)} \prod_{i=1}^d \varphi_i^{c_i-1} \left(1-\sum_{i=1}^d \varphi_i\right)^{(c_0-1)}, \tag{22}$$

where $\varphi_i \geq 0$ and $\sum_{i=1}^d \varphi_i \leq 1$. Panel (a) in Fig. 4 provides the level sets of the joint density of α_1 and β_1 with hyper-parameters $c_0 = 3$, $c_1 = 4$ and $c_2 = 3$. We assume a flat prior for α_0, i.e. $\pi(\alpha_0) = \mathbb{I}_{\mathbb{R}}(\alpha_0)$. For λ and ϕ we assume a joint prior distribution with uniform marginal prior $\lambda \sim \mathcal{U}_{[0,1]}$ and shifted gamma conditional prior $\phi \sim \mathcal{G}a^*(a,b,c)$, with density function:

$$\pi(\phi) = \frac{b^a}{\Gamma(a)}(\phi-c)^{(a-1)}e^{-b(\phi-c)} \quad \text{for } \phi > c, \tag{23}$$

where $c = (1-\lambda)^{-2}$. Panel (b) provides the level sets of the joint density function of ϕ and λ, with hyper-parameters $a = b = 5$. The joint prior distribution of the parameters will be denoted by $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\varphi})\pi(\alpha_0)\pi(\lambda)\pi(\phi)$.

4.2. Data augmentation

The probability distribution of Z_t is

$$f_t(Z_t = z|\boldsymbol{\theta}) = e^{-\sigma_t^2-z\lambda}\sum_{s=\underline{s}}^{+\infty}\frac{1}{4}\frac{\sigma_t^4+\mu_t^2}{s!(s+z)!}\left[\frac{\sigma_t^2+\mu_t}{2} + (s+z)\lambda\right]^{s+z-1}\left[\frac{\sigma_t^2-\mu_t}{2}+s\lambda\right]^{s-1}e^{-2\lambda s} \tag{24}$$

with $\underline{s} = \max(0,-z)$ and define $Z_{1:T} = (Z_1, \ldots, Z_T)$. Since the posterior distribution

$$\pi(\boldsymbol{\theta}|Z_{1:T}) \propto \prod_{t=1}^T f_t(Z_t|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{25}$$

is not analytically tractable, we apply Markov Chain Monte Carlo (MCMC) for posterior approximation in combination with a data-augmentation approach (Tanner & Wong, 1987). As in Karlis and Ntzoufras (2006), we exploit the stochastic representation of the GPD as the difference of GP latent variables $X_t \sim GP(\theta_{1t},\lambda)$ and $Y_t \sim GP(\theta_{2t},\lambda)$ with pmf $f_t(X_t = x|\theta_{1t},\lambda)$ and $f_t(Y_t = y|\theta_{2t},\lambda)$), respectively. Let $X_{1:T} = (X_1, \ldots, X_T)$ and $Y_{1:T} = (Y_1, \ldots, Y_T)$, then the complete-data likelihood becomes

$$f(Z_{1:T}, X_{1:T}Y_{1:T}|\boldsymbol{\theta}) = \prod_{t=1}^T \delta(Z_t - X_t + Y_t)f_t(X_t|\boldsymbol{\theta})f_t(X_t - Z_t|\boldsymbol{\theta}), \tag{26}$$

where $\delta(z-c)$ is the Dirac function that takes value 1 if $z = c$ and 0 otherwise. The joint posterior distribution of the parameters $\boldsymbol{\theta}$ and the latent variables $X_{1:T}$ and $Y_{1:T}$ is

$$\pi(X_{1:T}, Y_{1:T}, \boldsymbol{\theta}|Z_{1:T}) \propto f(Z_{1:T}, X_{1:T}Y_{1:T}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{27}$$

4.3. Gibbs sampler

We apply a Gibbs algorithm (Robert & Casella, 2013, Ch. 10) with a Metropolis–Hastings (MH) step. In the sampler, we draw the latent variables and the parameters of the model by iterating the following:

1. draw (X_t, Y_t) from $f(X_t, Y_t|Z_{1:T}, \boldsymbol{\theta})$ for $1 \leq t \leq T$;
2. draw $\boldsymbol{\varphi}$ from $\pi(\boldsymbol{\varphi}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\varphi})$;
3. draw ϕ from $\pi(\phi|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi})$;
4. draw λ from $\pi(\lambda|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda})$,

where $\boldsymbol{\theta}_{-\eta}$ indicates the collection of parameters excluding the element $\boldsymbol{\eta}$. The full conditional distribution of the latent variables satisfies

$$(X_t, Y_t) \sim f(Z_t|X_t, Y_t, \boldsymbol{\theta})f(X_t, Y_t|Z_{1:T}, \boldsymbol{\theta}). \tag{28}$$

We draw from the full conditional distribution by MH. Differently from Karlis and Ntzoufras (2006), we use a mixture proposal distribution, allowing for better MCMC chain mixing. At the jth iteration, we generate a candidate X_t^* from $GP(\theta_{1t},\lambda)$ with probability ν and $(X_t^* - Z_t)$

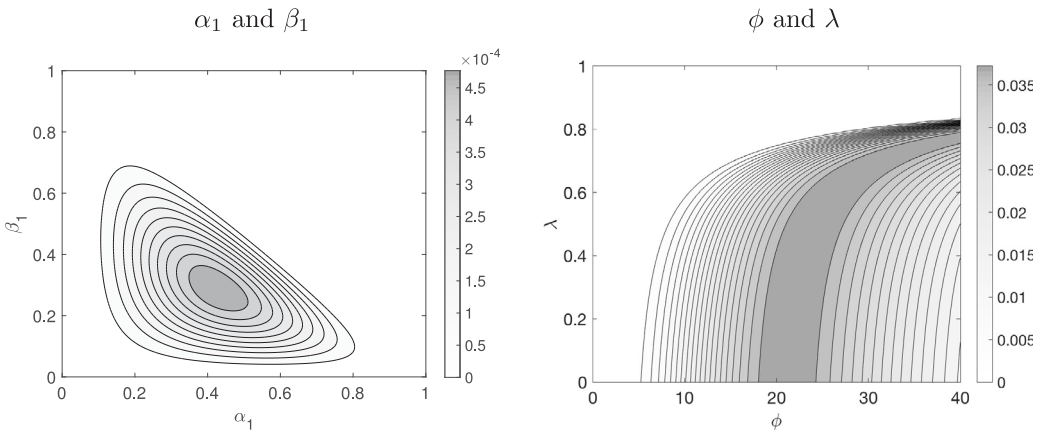

Fig. 4. Contour lines of the log-prior density function for α_1 and β_1 (left) and ϕ and λ (right).

from $GP(\theta_{2t}, \lambda)$ with probability $1-\nu$, and accept with probability

$$\varrho = \min\left\{1, \frac{f_t(X_t^*|\theta_{1t},\lambda)f_t(X_t^*-Z_t|\theta_{2t},\lambda)}{f_t(X_t^{(j-1)}|\theta_{1t},\lambda)f_t(X_t^{(j-1)}-Z_t|\theta_{2t},\lambda)}\frac{q(X_t^{(j-1)})}{q(X_t^*)}\right\} \tag{29}$$

where $q(X_t) = \nu f(X_t|\theta_{1t},\lambda) + (1-\nu)f(X_t - Z_t|\theta_{2t},\lambda)$ and $X_t^{(j-1)}$ is the $(j-1)$-th iteration value of the latent variable X_t.

Considering the parameter $\boldsymbol{\varphi}$, its full conditional density is

$$\boldsymbol{\varphi} \sim \pi(\boldsymbol{\varphi}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\varphi}) \propto \pi(\boldsymbol{\varphi})\prod_{t=1}^{T} f_t(X_t, Y_t|\boldsymbol{\theta}). \tag{30}$$

We adopt an MH step with Dirichlet independent proposal distribution $\boldsymbol{\varphi}^* \sim Dir(\mathbf{c}^*)$, where $\mathbf{c}^* = (c_0^*, c_1^*, c_2^*)$, and with acceptance probability

$$\varrho = \min\left\{1, \frac{\pi(\boldsymbol{\varphi}^*|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\varphi})}{\pi(\boldsymbol{\varphi}^{(j-1)}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\varphi})}\right\}. \tag{31}$$

The full conditional density of ϕ is

$$\pi(\phi|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi}) \propto \pi(\phi)\prod_{t=1}^{T} f_t(X_t, Y_t|\boldsymbol{\theta}). \tag{32}$$

We consider the change of variable $\zeta = \log(\phi - c)$ with Jacobian $\exp(\zeta)$ and a MH step with a random walk proposal $\zeta^* \sim N(\zeta^{(j-1)}, \gamma^2)$, where $\zeta^{(j-1)} = \log(\phi^{(j-1)} - c)$, $\phi^{(j-1)}$ is the previous value of the parameter, and $c = 1/(1-\lambda)^2$. The acceptance probability is

$$\varrho = \min\left\{1, \frac{\pi(\phi^*|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi})\exp(\zeta^*)}{\pi(\phi^{(j-1)}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi})\exp(\zeta^{(j-1)})}\right\}, \tag{33}$$

where $\phi^* = c + \exp(\zeta^*)$.

Samples from the full conditional density of λ

$$\pi(\lambda|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda}) \propto \pi(\lambda)\prod_{t=1}^{T} f_t(X_t, Y_t|\boldsymbol{\theta}). \tag{34}$$

are obtained by a MH with Beta random walk proposal $\lambda^* \sim Be(s\lambda^{(j-1)}, s(1-\lambda^{(j-1)}))$, where s is a precision parameter. The acceptance probability is:

$$\varrho = \min\left\{1, \frac{\pi(\lambda^*|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda})Be(s\lambda^*, s(1-\lambda^*))}{\pi(\lambda^{(j-1)}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda})Be(s\lambda^{(j-1)}, s(1-\lambda^{(j-1)}))}\right\}. \tag{35}$$

The MCMC samples $\boldsymbol{\theta}^{(j)}, j = 1, \ldots, J$ are used to evaluate the approximated Bayes estimator

$$\hat{\boldsymbol{\theta}} = \frac{1}{J}\sum_{j=1}^{J}\boldsymbol{\theta}^{(j)}.$$

4.4. Forecasting

The Gibbs sampler above can approximate point and distribution forecasts of the variables of interest Z_{T+h}, $h = 1, \ldots, H$, where H is the forecasting horizon. At the jth iteration, we draw $Z_{T+h}^{(j)}$ from the conditional distribution given past observations $Z_1, \ldots, Z_T$, and the parameter draws $\boldsymbol{\theta}^{(j)}$, as

$$Z_{T+h}^{(j)}|\mathcal{F}_T, \boldsymbol{\theta}^{(j)} \sim GPD\left(\mu_{T+h}^{(j)}, \sigma_{T+h}^{2,(j)}, \lambda^{(j)}\right), \tag{36}$$

where $j = 1, \ldots, J$, denotes the MCMC draw and

$$\mu_{T+h}^{(j)} = \begin{cases} \alpha_0^{(j)}(1-\lambda^{(j)}) + \alpha_1^{(j)}(1-\lambda^{(j)})Z_T \\ +\beta_1^{(j)}\mu_T^{(j)}, & \text{for } h = 1 \\ \alpha_0^{(j)}(1-\lambda^{(j)}) + \alpha_1^{(j)}(1-\lambda^{(j)})Z_{T+h-1}^{(j)} \\ +\beta_1^{(j)}\mu_{T+h-1}^{(j)}, & \text{for } h = 2, \ldots, H \end{cases} \tag{37}$$

$$\sigma_{T+h}^{2(j)} = |\mu_{T+h}^{(j)}|\phi^{(j)}(1-\lambda^{(j)})^2. \tag{38}$$

Forecasts for the variables in the level $V_{T+h} = V_{T+h-1} + Z_{T+h}$ can be easily obtained from the recursion

$$V_{T+h}^{(j)} = V_{T+h-1} + Z_{T+h}^{(j)}, \quad h = 1, \ldots, H, \tag{39}$$

with initial value $V_T^{(j)}$ equal to the sample V_T available at time T, where the GPD increments $Z_{T+h}^{(j)}$ are sampled from $GPD\left(\mu_{T+h}^{(j)}, \sigma_{T+h}^{2,(j)}, \lambda^{(j)}\right)$ under the constraint: $Z_{T+h}^{(j)} \geq$

Table 1
Autocorrelation function (ACF), effective sample size (ESS), and inefficiency factor (INEFF) of the posterior MCMC samples for the settings: low persistence and high persistence. The results are averages over 50 independent MCMC experiments on 50 independent datasets of 400 observations each. We ran the proposed MCMC algorithm for 1,010,000 iterations and evaluated the statistics before (subscript BT) and after (subscript AT), removing the first 10,000 burn-in samples and applying a thinning procedure with a factor of 250. The p-values of Geweke's convergence diagnostic are provided between parentheses.

	Low persistence ($\alpha = 0.25$, $\beta = 0.23$, $\lambda = 0.4$)			High persistence ($\alpha = 0.53$, $\beta = 0.25$, $\lambda = 0.6$)		
	α	β	λ	α	β	λ
$ACF(1)_{BT}$	0.96	0.97	0.97	0.91	0.88	0.98
$ACF(10)_{BT}$	0.86	0.83	0.81	0.70	0.52	0.83
$ACF(30)_{BT}$	0.75	0.69	0.63	0.52	0.37	0.60
$ACF(1)_{AT}$	0.43	0.39	0.27	0.21	0.13	0.16
$ACF(10)_{AT}$	0.25	0.18	0.12	0.20	0.06	0.11
$ACF(30)_{AT}$	0.18	0.15	0.07	0.15	0.06	0.09
ESS_{BT}	0.02	0.02	0.02	0.02	0.03	0.02
ESS_{AT}	0.07	0.07	0.09	0.09	0.12	0.11
$INEFF_{BT}$	50.53	51.07	43.88	48.39	43.35	49.25
$INEFF_{AT}$	26.36	27.29	13.99	17.21	16.84	12.59
CD_{BT}	11.81	−28.69	0.78	0.93	−6.27	2.40
	(0.11)	(0.14)	(0.10)	(0.04)	(0.06)	(0.05)
CD_{AT}	5.72	−13.18	0.2	0.74	−3.84	1.17
	(0.23)	(0.23)	(0.23)	(0.13)	(0.15)	(0.11)

$V_{T+h-1}^{(j)}$. The point forecast $\mathbb{E}(V_{T+h}|\mathcal{F}_T)$ can be approximated as follows

$$\widehat{\mathbb{E}(V_{T+h}|\mathcal{F}_T)} = \frac{1}{J}\sum_{j=1}^{J} V_{T+h}^{(j)} \tag{40}$$

and, similarly, other quantities of interest, such as predictive distribution and quantiles, can be approximated using the simulated values $Z_{T+h}^{(j)}$.

5. Simulation study

The objectives of our simulation experiments are manifold: firstly, we test for the correct implementation of this algorithm presented in Section 4; secondly, to assess the efficiency of the algorithm; thirdly, to explore the sensitivity of the results to the prior specification; and finally, to demonstrate the impact of model misspecification on the GARCH coefficient estimates.

We apply Geweke's (2004) procedure to test for the correct implementation of the Gibbs sampler. We used 2,000 MCMC samples and the first three moments as test functions for each parameter. Our Gibbs sampler withstands all tests. The values of the statistics are reported in Appendix B. See Section S.2 in Supplement Material for background material on Geweke's test. As regards the efficiency of the MCMC, we computed the (Geweke, 1992) convergence diagnostic measure (CD), the inefficiency factor (INEFF)[1] and the Effective Sample Size (ESS).

We simulated 50 independent series of 400 observations each. We ran the Gibbs sampler for 1,010,000 iterations on each dataset, discarded the first 10,000 draws to remove dependence on initial conditions and applied a thinning procedure with a factor of 250 to reduce the dependency between consecutive draws. For each setting, Appendix B reports the output of one of the experiments. Since in the applications, the specification with $p = q = 1$ provides the best fitting, in the following, we present the results for this case and define $\alpha_1 = \alpha$ and $\beta_1 = \beta$, for ease of notation.

As in GARCH and stochastic volatility modelling (see, e.g. Chib, Nardari, and Shephard, 2002 and references therein), we test the efficiency of the algorithm in two different settings: low persistence and high persistence. The true values of the parameters are: $\alpha = 0.25$, $\beta = 0.23$, $\lambda = 0.4$ in the low persistence setting and $\alpha = 0.53$, $\beta = 0.25$, $\lambda = 0.6$ in the high persistence setting. Table 1 shows, for the parameters α, β, and λ, the INEFF, ESS, and ACF averaged over the 50 replications before (BT subscript) and after thinning (AT subscript).

The thinning procedure reduces the autocorrelation levels and increases the ESS, especially in the high persistence setting. The p-values of the CD statistics indicate that the null hypothesis that two sub-samples of the MCMC draws have the same distribution is accepted. The efficiency of the MCMC algorithm generally improved after thinning. On average, the inefficiency measures (19.05), the p-values of the CD statistics (0.18), and the acceptance rates (0.35) achieve the values recommended in the literature (e.g., see Roberts, Gelman, Gilks, et al., 1997).

We investigate the impact on inference of a prior distribution favouring nonstationarity, that is, a prior with larger probability mass at values of $\alpha + \beta$ close to 1. The results in Figs. B.8 and B.9 in Appendix B show that in both high and low persistence settings, the information from

[1] The inefficiency factor is defined as

$$INEFF = 1 + 2\sum_{k=1}^{\infty}\rho(k)$$

where $\rho(k)$ is the sample autocorrelation at lag k for the parameter of interest. It measures how well the MCMC chain mixes. An INEFF equal to m tells us that one needs to draw MCMC samples m times as many as uncorrelated samples.

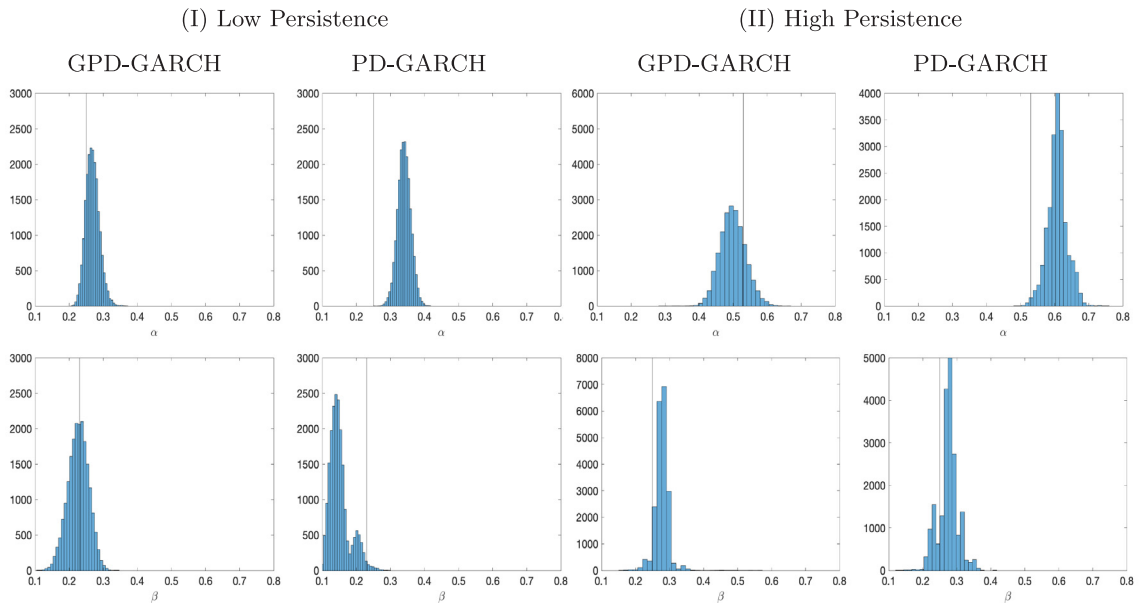

Fig. 5. Posterior histograms for α (first row) and β (second row) in the low persistence case (left panel) and high persistence case (right panel), when we fit the correct model $GPD-INGARCH$ and the misspecified model $PD-INGARCH$.

the data dominates the posterior, which is not particularly sensitive to the prior choice.

Finally, we study estimation bias when the model is not correctly specified. Since the GARCH dynamics in μ_t is related to the over-dispersion λ and the conditional mean of the process (see Eq. (12)), one can expect some estimation bias when λ is not correctly specified. In our experiments, when a PD-INGARCH is estimated on data generated from a GPD-INGARCH, the GARCH parameters are biased. For illustrative purposes, we report in Fig. 5 the results of some of our experiments. The estimation bias for α and β is 0.11 and -0.09, respectively, in the low persistence case and 0.1 and 0.02, respectively, in the high persistence.

6. Real data examples

6.1. Car accident data

The dataset for this application consists of the number of car accidents near Schiphol airport in The Netherlands during 2001, sampled at a daily frequency (see Fig. 6). This set of 365 observations has been previously examined in Andersson and Karlis (2014) and Brijs et al. (2008). Based on the results of the Augmented Dickey–Fuller and Phillips–Perron tests for unit roots, the time series of accident counts is found to be non-stationary. In contrast, the first differences do not exhibit unit roots (refer to Table C.1 in Appendix C). Therefore, we applied the estimation procedure outlined in Section 4 to the first differences. The Gibbs sampler was run for 110,000 iterations, with the first 10,000 discarded as a burn-in sample. The final sample was then thinned down at a rate of 10%.

Table 2 presents the parameter posterior mean and standard error and the 95% credible interval for the unrestricted INGARCH(1,1) model (model $\mathcal{M}_1$). See Fig. C.1 in the Appendix for the histograms of posteriors. There is evidence of persistence in the expected accident arrivals, $\hat{\alpha}+\hat{\beta}=0.203$, heteroskedastic effects, $\hat{\beta}=0.161$, skewness $\hat{\lambda}=0.023$, and overdispersion $\hat{\phi}=17.203$.

We study the contribution of overdispersion, heteroskedasticity, and persistence by testing some restrictions of the INGARCH(1,1). Model $\mathcal{M}_2$ corresponds to an INGARCH(1, 1) where the observations are from the standard Poisson-difference GARCH model PD-INGARCH(1, 1). Model $\mathcal{M}_3$ corresponds to an autoregressive model, GPD-INARCH(1, 0), whereas $\mathcal{M}_4$ is a standard Poisson difference autoregressive model, PD-INARCH(1, 0). The estimates of the restricted models $\mathcal{M}_2$ to $\mathcal{M}_4$ are given in Table 2.

In Bayesian analysis, model comparison can be carried out in different ways, such as through the Deviance Information Criterion (DIC) and Bayes Factor (BF). Using DIC for latent variable models presents difficulties, as discussed in Celeux, Forbes, Robert, and Titterington (2006). Thus, we compare models via BF, which is the ratio of normalizing constants of the posterior distributions of two potential models (see Cameron and Pettitt (2014) for a review). MCMC methods allow for generating samples from the posterior distributions, which can then be exploited to estimate the ratio of normalizing constants. In this paper, we follow the RLR method proposed by Geyer (1994). As acknowledged by Gelman and Meng (1998), RLR is a specific form of bridge sampling. It provides a very efficient tool for BF approximation since logistic regression is readily available in standard statistical software. We refer the reader to Llorente, Martino, Delgado, and Lopez-Santiago (2023) for a recent and authoritative

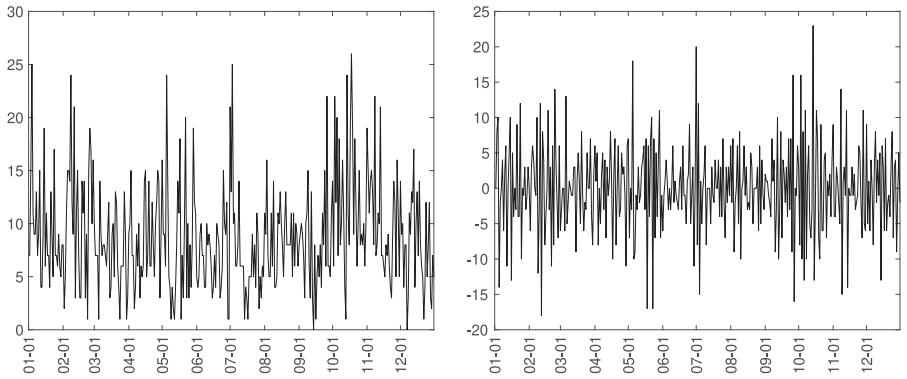

Fig. 6. Frequency (left) and day-on-day changes (right) of the accidents at the Schiphol airport in The Netherlands in 2001.

Table 2
Model estimates for the Schiphol's dataset. Posterior mean (Mean), standard deviation (Std) and 95% credible intervals (CI) of the parameters θ are provided for different specifications of the GPD-INGARCH (different panels), where $\mathcal{M}_1$ is the GPD-INGARCH(1,1), $\mathcal{M}_2$ is the PD-INGARCH(1,1) with $\lambda = 0$, $\mathcal{M}_3$ is the GPD-INARCH(1,0) and $\mathcal{M}_4$ is the PD-INARCH(1,0) with $\lambda = 0$.

θ	Mean	Std	CI	Mean	Std	CI
	$\mathcal{M}_1$: GPD-INGARCH(1,1)			$\mathcal{M}_2$: PD-INGARCH(1,1) and $\lambda = 0$		
α_0	−0.051	0.017	(−0.088, −0.025)	−0.051	0.016	(−0.08, −0.023)
α	0.042	0.014	(0.021, 0.073)	0.042	0.013	(0.019, 0.066)
β	0.161	0.012	(0.142, 0.189)	0.158	0.011	(0.141, 0.189)
λ	0.023	0.01	(0.007, 0.044)	–	–	–
ϕ	17.203	5.68	(9.001, 30.708)	16.67	6.479	(9.495, 32.181)
	$\mathcal{M}_3$: GPD-INARCH(1,0)			$\mathcal{M}_4$: PD-INARCH(1,0) and $\lambda = 0$		
α_0	−0.063	0.024	(−0.118, −0.024)	−0.066	0.02	(−0.113, −0.031)
α	0.039	0.014	(0.015, 0.069)	0.041	0.013	(0.02, 0.071)
β	–	–	–	–	–	–
λ	0.02	0.01	(0.005, 0.042)	–	–	–
ϕ	18.547	8.404	(8.864, 41.224)	15.814	5.277	(8.428, 29.907)

survey of marginal likelihood approximation methods in Bayesian hypothesis testing. RLR consists of deriving the normalizing constants by logistic regression, considering both samples as if they were issued from a mixture of two distributions with probability

$$p_j(v, \eta_j) = \frac{h_j(v)\exp(\eta_j)}{h_1(v)\exp(\eta_1) + h_2(v)\exp(\eta_2)}, \; j = 1, 2 \qquad (41)$$

to be generated from the jth distribution of the mixture. Geyer (1994) proposes to estimate the log-Bayes factor $\kappa = \eta_2 - \eta_1$ by maximizing the quasi-likelihood function

$$\ell_n(\kappa) = \sum_{i=1}^{n} \log p_1(V_{i1}, \eta_1) + \sum_{i=1}^{n} \log p_2(V_{i2}, \eta_2), \qquad (42)$$

where n is the number of MCMC draws for each model and $V_{ij} = \log f_j(Z_{1:T}, X_{j,1:T}^{(i)}, Y_{j,1:T}^{(i)} | \boldsymbol{\theta}_j^{(i)})$ is the log-likelihood of the model j evaluated at the ith MCMC sample.

We ran six reverse logistic regressions and performed pairwise comparisons of our models. The approximated logarithmic Bayes factors $BF(\mathcal{M}_i, \mathcal{M}_j)$ are given in Table 3. On the entire sample, our GPD-INGARCH(1, 0), that is, the GPD-INARCH, overperforms the other models (bottom-right panel). However, the stability of a prediction model is a crucial factor in ensuring accurate forecasts. Consequently, we assessed the in-sample and out-of-sample predictive performances of different model specifications. Rolling estimation and prediction exercises were conducted from June 2001 to December 2001, utilizing windows of 160 observations. We evaluated the in-sample Bayes factors between model pairs for each window and computed the h-step-ahead posterior predictive distribution for the number of accidents. We followed the procedure outlined in Section 4, which enables the incorporation of parameter uncertainty into the forecasts (see, e.g., McCabe & Martin, 2005; McCabe et al., 2011, and references therein).

Given the Mean Square Forecasting Error (MSFE) of each model, we computed in Fig. 7 the model weights defined as $w_j = \exp(-\text{MSFE}(\mathcal{M}_j))/(\exp(-\text{MSFE}(\mathcal{M}_1)) + \cdots + \exp(-\text{MSFE}(\mathcal{M}_4)))$ (Billio, Casarin, Ravazzolo, & Van Dijk, 2013). There is evidence of superior performance by the unrestricted GPD-GARCH during periods characterized by larger variability and overdispersion. For illustrative purposes, we have presented the BFs of the in-sample analyses for three periods: January 23, 2001, to March 14, 2001; May 4, 2001, to June 23, 2001; and November 11, 2001, to December 31, 2001 (see Table 3). To streamline the presentation, we have reported results at horizons $h = 1, 3, 5, 7$. In the out-of-sample analysis, GPD models $\mathcal{M}_1$ and $\mathcal{M}_3$ (black lines) outperform others at all horizons during periods characterized by higher overdispersion,

Table 3
In-sample logarithmic Bayes Factors for the Schiphol's dataset, BF($\mathcal{M}_i$, $\mathcal{M}_j$), of the model $\mathcal{M}_i$ (rows) against model $\mathcal{M}_j$ (columns), with $i < j$, where $\mathcal{M}_1$ is the GPD-INGARCH(1,1), $\mathcal{M}_2$ is the PD-INGARCH(1,1) with $\lambda = 0$, $\mathcal{M}_3$ is the GPD-INARCH(1,0) and $\mathcal{M}_4$ is the PD-INARCH(1,0) with $\lambda = 0$. Numbers in parentheses are Monte Carlo standard deviations of the estimated Bayes factors. Bayes factors are evaluated on the whole sample and sub-samples of 50 periods (different panels).

BF($\mathcal{M}_i$, $\mathcal{M}_j$)	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
	23 Jan 2001–14 Mar 2001			04 May 2001–23 Jun 2001		
$\mathcal{M}_1$	95.99	108.79	466.10	190.04	600.33	835.24
	(0.56)	(0.63)	(2.72)	(1.10)	(3.49)	(4.86)
$\mathcal{M}_2$		−41.58	91.81		475.67	707.74
		(0.24)	(0.54)		(2.77)	(4.12)
$\mathcal{M}_3$			427.36			111.04
			(2.50)			(0.65)
	11 Nov 2001–31 Dec 2001			Full sample (01 Jan 2001–31 Dec 2001)		
$\mathcal{M}_1$	70.84	429.97	2.06e+03	278.25	−3.46e+03	−3.89e+03
	(0.42)	(2.59)	(12.46)	(0.22)	(2.71)	(3.04)
$\mathcal{M}_2$		465.86	1.40e+03		−6.15e+03	−7.07e+03
		(2.81)	(8.49)		(4.81)	(5.53)
$\mathcal{M}_3$			576.21			1.24e+03
			(3.51)			(0.97)

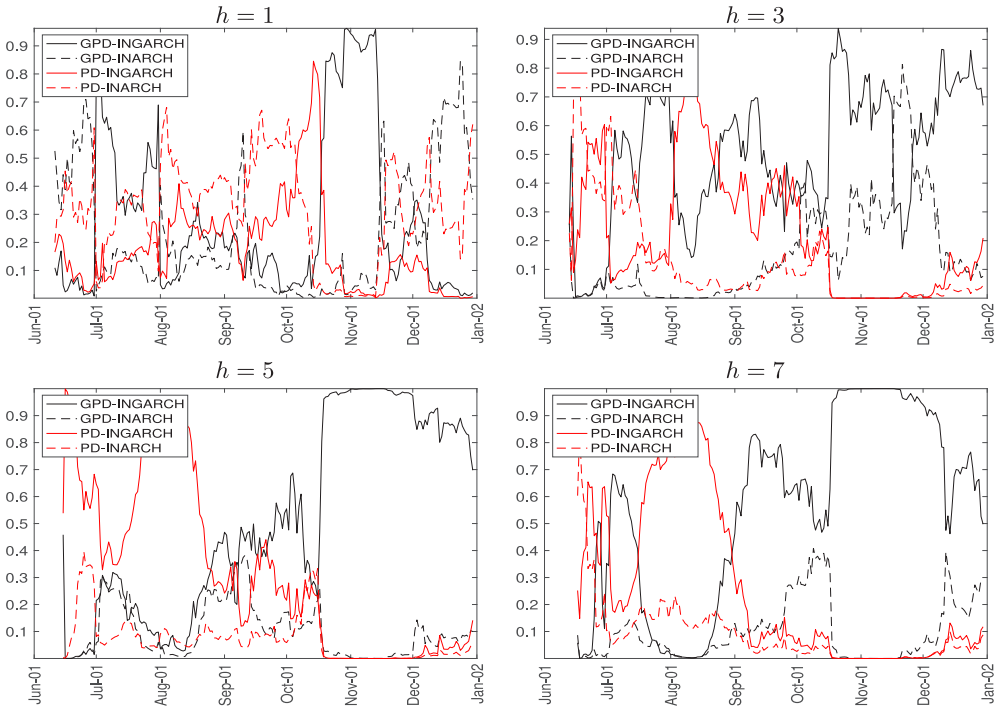

Fig. 7. Rolling out-of-sample predictive performances for the Schiphol's dataset. Given the Mean Square Forecasting Errors (MSFE) of all models under consideration ($\mathcal{M}_j$ $j = 1, \ldots, 4$), the model weight is computed as $w_j = \exp(-\text{MSFE}(\mathcal{M}_j))/(\exp(-\text{MSFE}(\mathcal{M}_1)) + \cdots + \exp(-\text{MSFE}(\mathcal{M}_4)))$, for different horizons h (different plots) and models $\mathcal{M}_j$ (different lines and colours).

such as November-December 2001. However, at longer horizons, exceeding two weeks, performances become comparable.

6.2. Cyber threat data

According to the Financial Stability Board (FSB, 2018, pp. 8–9), a cyber incident is any observable occurrence in an information system that jeopardizes the cyber security of the system or violates the security policies and procedures or the use policies. Over the past years, there have been several discussions on the taxonomy of incidents classification (see, e.g. ENISA, 2018). This paper uses the classification in the Passeri (2019) dataset. Passeri (2019) is a cyber-incident website that collects public reports and provides the number of cyber incidents for different categories of threats: crimes, espionage and warfare. Fig. 8 shows the total and category-specific number of cyber attacks at a daily frequency from January 2017 to December

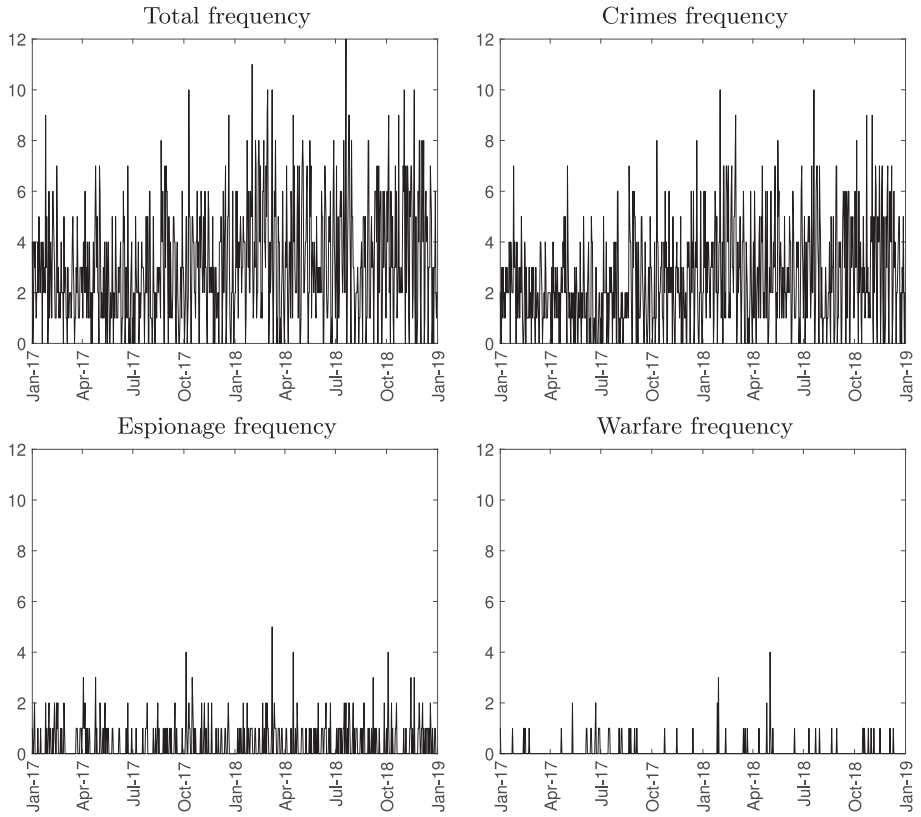

Fig. 8. Daily cyber-threats counts between 1st January 2017 and 31st December 2018.

2018. Albeit limited in its variety of cyber attacks, the dataset covers some relevant cyber events and is one of the few publicly available datasets (Agrafiotis et al., 2018). The daily threat frequencies are between 0 and 12, which motivates using a discrete distribution. We remove the stochastic trend by considering the first difference and then fit the GPD-INGARCH model.

We applied our estimation procedure of Section 4 and ran the Gibbs sampler for 110,000 iterations, discarded the first 10,000 as a burn-in sample, and thinned down the sample with a rate of 10%. Our results in Panel $\mathcal{M}_1$ of Table 4 suggest that the GPD-INGARCH is a suitable model for the cyber threat dataset as it effectively captures both persistence and heteroskedastic effects, indicated by $\hat{\alpha} + \hat{\beta} = 0.395$ and $\hat{\beta} = 0.333$. Moreover, our estimates $\hat{\lambda} = 0.327$ and $\hat{\phi} = 92.155$ indicate the presence of skewness and over-dispersion in the data. Fig. C.2 presents the corresponding posterior histograms.

We investigate the heteroskedastic and overdispersion features of the data by comparing different models via BF. On the whole sample, the PD-INGARCH(1,1) model overperforms the other, providing evidence of overdispersion $\hat{\phi} = 72.231$ and confirming the absence of skewness in the conditional distribution, i.e., $\lambda = 0$ (see Table 4 and the bottom-right panel of Table 5). The estimates $\hat{\beta} = 0.333$ and $\hat{\beta} = 0.388$ for $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively, and the fact that both PD- and GPD-INGARCH have better fitting than PD- and GPD-INARCH indicates heteroskedastic effects in the data.

We study the in-sample and the out-of-sample predictive performances of different model specifications. Rolling estimation and prediction exercises are considered from January 2018 to December 2018, with windows of 365 observations. We evaluate the BF between model pairs and the model weights based on each model's MSFE, as summarized in Table 5. We reported the model weights at some horizons, i.e. $h = 1, 3, 5, 7$, for illustrative purposes. In our results, GPD-INGARCH(1,1) overperforms other models during periods with a larger number of threats and higher overdispersion. For the in-sample model comparison, see, e.g., the results for February-March 2018 and May-July 2018 (top panels in Table 5). This result is confirmed by the out-of-sample forecasting, especially at larger horizons, 5 and 7 days ahead (solid black lines in panels $h = 5$ and $h = 7$ of Fig. 9). The GPD model class can better capture spikes in the series (Fig. C.5 in the Appendix). When the number of attacks, their conditional skewness, and overdispersion are lower, PD-INGARCH(1,1) provides better results in and out of the sample. The evidence points to variations in the

Table 4
Model estimates for the cyber treat dataset. Posterior mean (Mean), standard deviation (Std) and 95% credible intervals (CI) of the parameters θ are provided for different specifications of the GPD-INGARCH (different panels), where $\mathcal{M}_1$ is the GPD-INGARCH(1,1), $\mathcal{M}_2$ is the PD-INGARCH(1,1) with $\lambda = 0$, $\mathcal{M}_3$ is the GPD-INARCH(1,0) and $\mathcal{M}_4$ is the PD-INARCH(1,0) with $\lambda = 0$.

θ	Mean	Std	CI	Mean	Std	CI
	$\mathcal{M}_1$: GPD-INGARCH(1,1)			$\mathcal{M}_2$: PD-INGARCH(1,1) and $\lambda = 0$		
α_0	−0.057	0.011	(−0.08, −0.039)	−0.041	0.005	(−0.051, −0.034)
α	0.062	0.012	(0.043, 0.089)	0.055	0.009	(0.045, 0.086)
β	0.333	0.014	(0.307, 0.351)	0.388	0.029	(0.295, 0.403)
λ	0.327	0.013	(0.303, 0.354)	–	–	–
ϕ	92.155	17.117	(64.241, 127.834)	64.681	6.989	(48.230, 77.582)
	$\mathcal{M}_3$: GPD-INARCH(1,0)			$\mathcal{M}_4$: PD-INARCH(1,0) and $\lambda = 0$		
α_0	−0.042	0.007	(−0.057, −0.031)	0.021	0.004	(0.014, 0.027)
α	0.07	0.011	(0.052, 0.094)	0.044	0.008	(0.03, 0.058)
β	–	–	–	–	–	–
λ	0.333	0.013	(0.309, 0.358)	–	–	–
ϕ	83.601	12.715	(61.275, 111.008)	72.231	13.7568	(54.505, 105.217)

Table 5
In-sample logarithmic Bayes Factors for the cyber threat dataset, BF($\mathcal{M}_i$, $\mathcal{M}_j$), of the model $\mathcal{M}_i$ (rows) against the model $\mathcal{M}_j$ (columns), with $i < j$, where $\mathcal{M}_1$ is the GPD-INGARCH(1,1), $\mathcal{M}_2$ is the PD-INGARCH(1,1) with $\lambda = 0$, $\mathcal{M}_3$ is the GPD-INARCH(1,0) and $\mathcal{M}_4$ is the PD-INARCH(1,0) with $\lambda = 0$. Numbers in parentheses are Monte Carlo standard deviations of the estimated Bayes factors. Bayes factors are evaluated on the whole sample and sub-samples of 50 periods (different panels).

BF($\mathcal{M}_i$, $\mathcal{M}_j$)	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
	05 Feb 2018–27 Mar 2018			26 May 2018–15 Jul 2018		
$\mathcal{M}_1$	73.30 (0.32)	114.52 (0.51)	18.18 (0.08)	44.68 (0.22)	424.87 (2.19)	27.14 (0.14)
$\mathcal{M}_2$		102.25 (0.46)	−90.22 (0.39)		496.43 (2.57)	−68.43 (0.34)
$\mathcal{M}_3$			−329.19 (1.44)			−769.36 (3.97)
	11 Nov 2018–31 Dec 2018			Full sample (02 Jan 2017–31 Dec 2018)		
$\mathcal{M}_1$	−1.63e+03 (6.47)	18.30 (0.08)	−182.37 (0.76)	−3.18e+03 (1.04)	488.62 (0.17)	−3.54e+03 (1.16)
$\mathcal{M}_2$		1.47e+03 (5.87)	2.11e+03 (8.28)		2.47e+03 (0.82)	406.80 (0.13)
$\mathcal{M}_3$			−456.23 (1.93)			−2.69e+03 (0.89)

performances of two model classes: ARCH and GARCH. At the one-day-ahead horizon, INARCH models often over-perform INGARCH models, whereas at larger horizons when the heteroscedastic effects are stronger (e.g., June-August 2018 in panel $h = 5$), both PD- and GPD-INGARCH models (solid lines in Fig. 9) perform better than INARCH models (dashed lines). Our findings naturally call for dynamic and horizon-specific model selection or combination strategies (Billio et al., 2013; McAlinn & West, 2019), which we leave for future research.

7. Conclusions

A new family of stochastic processes with values in the set of signed integers has been introduced in this paper, where increments of the process follow a generalized Poisson difference distribution, and the parameters have GARCH dynamics. Relevant properties have been derived via a thinning representation of the process, and a consistent forecasting procedure has been derived. We adopted a Bayesian inference framework and developed an efficient Monte Carlo Markov Chain sampler to approximate the posterior distribution and incorporate parameter uncertainty in the predictive distribution. Our simulation studies show that the MCMC algorithm is efficient in different parameter settings and that inference results are robust with respect to the prior specification. The findings suggest that standard Poisson difference models yield biased estimates of the GARCH dynamics when observations are generated from a generalized Poisson model with large enough overdispersion. This modelling approach is validated by analysing a benchmark dataset on car accidents and an original application to cyber-threat data. The in-sample and out-of-sample model comparisons reveal that the proposed GPD-INGARCH model best captures over-dispersion and persistence in conditional moments. Notably, the forecasting abilities of the models exhibit temporal variations and differ across forecasting horizons for both datasets.

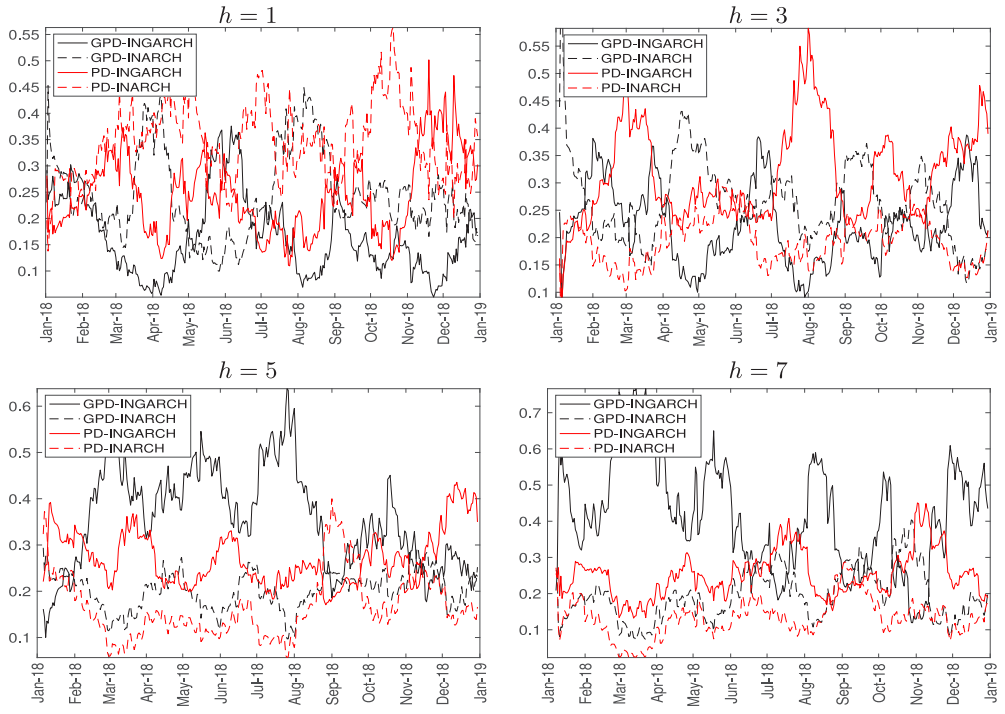

Fig. 9. Rolling out-of-sample predictive performances for the cyber threat dataset. Given the Mean Square Forecasting Errors (MSFE) of all models under consideration ($\mathcal{M}_j$ $j = 1, \ldots, 4$), the model weight is computed as $w_j = \exp(-\text{MSFE}(\mathcal{M}_j))/(\exp(-\text{MSFE}(\mathcal{M}_1)) + \cdots + \exp(-\text{MSFE}(\mathcal{M}_4)))$, for different horizons h (different plots) and models $\mathcal{M}_j$ (different lines and colours).

These observations naturally prompt further exploration of dynamic and horizon-specific model selection and combination for generalized Poisson models.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors' research used the SCSCF and HPC multiprocessor cluster systems at the Ca' Foscari University of Venice. This work was funded in part by the Université Franco-Italienne '*Visiting Professor Grant*' UIF 2017–VP17_12, by the French government under the management of Agence Nationale de la Recherche as part of the '*Investissements d'avenir*' program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute); the Institut Universitaire de France; the MUR – PRIN project '*Discrete random structures for Bayesian learning and prediction*' under g.a. n. 2022CLTYP4 and the Next Generation EU – '*GRINS – Growing Resilient, INclusive and Sustainable*' project (PE0000018), National Recovery and Resilience Plan (NRRP) – PE9. The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Code and datasets

The author(s) published code and data associated with this article on Code Ocean, a computational reproducibility platform. See: https://doi.org/10.24433/CO.1787135.v1.

Appendix A. Proof of the results contained in the paper

A.1. Proofs of the results in Section 2

We state the following preliminary results before proving the results in Lemma 1.

Lemma A.1. *Let $X \sim GP(\theta_1, \lambda)$ and $Y \sim GP(\theta_2, \lambda)$ be two independent GP random variables. The probability distribution of $Z = (X - Y)$ follows a Generalized Poisson Difference distribution (GPD) with pmf:*

$$P_z(\theta_1, \theta_2, \lambda) = e^{-\theta_1-\theta_2-z\lambda} \sum_{y=\max(0,-z)}^{\infty} \frac{\theta_2(\theta_2+y\lambda)^{y-1}}{y!} \frac{\theta_1(\theta_1+(y+z)\lambda)^{y+z-1}}{(y+z)!} e^{-2y\lambda}, \tag{A.1}$$

where z takes values in $\mathbb{Z}$, $-1 < \lambda < 1$ and $\theta_1, \theta_2 > 0$ are the parameters of the distribution. We denote this distribution as $GPD(\theta_1, \theta_2, \lambda)$.

Proof. See Consul (1986). □

In this paper, we used an alternative parametrization of the GPD that is better suited for modelling purposes. The following lemma shows that our parametrization is equivalent to the one used in Consul (1986).

Lemma A.2. *The probability distribution in Eq. (2) is equivalent to the one in Eq. (A.1) via the reparametrization* $\mu = \theta_1 - \theta_2$ *and* $\sigma^2 = \theta_1 + \theta_2$.

Proof of Lemma A.2. Let $Z \sim GPD(\theta_1, \theta_2, \lambda)$. We introduce the new parametrization: $\theta_1 = (\sigma^2 + \mu)/2$ and $\theta_2 = (\sigma^2 - \mu)/2$. By substituting θ_1 and θ_2 into Eq. (A.1) we have

$$\mathbb{P}(\{Z = z\}) = e^{\sigma^2 - z\lambda} \sum_{y=\max(0,-z)}^{+\infty} \frac{\sigma^4 - \mu^2}{2y!(y+z)!} \left(\frac{\sigma^2 + \mu}{2} + (y+z)\lambda\right)^{y+z-1} \left(\frac{\sigma^2 - \mu}{2} + y\lambda\right)^{y-1} e^{-2\lambda y}. \quad \square \tag{A.2}$$

Proof of Lemma 1. If $Z \sim GPD(\theta_1, \theta_2, \lambda)$ in the parametrization of Consul (1986), the moments are given in the Supplementary Material. By using our reparametrization of the GPD $\theta_1 = (\sigma^2 + \mu)/2$ and $\theta_2 = (\sigma^2 - \mu)/2$ we obtain mean, variance, skewness, and kurtosis:

$$\begin{aligned} \mathbb{E}(Z) &= \frac{\theta_1 - \theta_2}{1-\lambda} = \frac{\mu}{1-\lambda}, \\ \mathbb{V}(Z) &= \frac{\theta_1 + \theta_2}{(1-\lambda)^3} = \frac{\sigma^2}{(1-\lambda)^3} = \kappa_2 \\ S(Z) &= \frac{\mu^{(3)}}{\sigma^3} = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{\mu}{\sigma^3}\frac{(1+2\lambda)}{\sqrt{1-\lambda}}, \\ K(Z) &= \frac{\kappa_4 + 3\kappa_2^2}{\sigma^4} = 3 + \frac{1 + 8\lambda + 6\lambda^2}{\sigma^2(1-\lambda)}, \end{aligned} \tag{A.3}$$

where κ_i, $i = 2, 3, 4$ are respectively the second, third and fourth cumulants. $\square$

Proof of Proposition 1. Let ψ_j be the coefficient of z^j in the Taylor expansion of $G(z)D(z)^{-1}$. We have

$$\begin{aligned} \mu &= \mathbb{E}(Z_t) = \mathbb{E}(\mathbb{E}(Z_t|\mathcal{F}_{t-1})) = \mathbb{E}\left(\alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j Z_{t-j}\right) \\ &= \alpha_0 D^{-1}(1) + \mu D^{-1}(1) G(1) \qquad \text{(A.4)} \\ \Leftrightarrow \mu &= \frac{\alpha_0}{D(1) - G(1)} = \alpha_0 \left(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j\right)^{-1} \\ &= \alpha_0 K^{-1}(1), \qquad \text{(A.5)} \end{aligned}$$

where the equality before last in Eq. (A.4) follows from

$$\begin{aligned} \mathbb{E}(Z_t|\mathcal{F}_{t-1}) &= \mathbb{E}(\alpha_0 D^{-1}(1) + H(B) Z_t|\mathcal{F}_{t-1}) \\ &= \alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j Z_{t-j}. \end{aligned} \tag{A.6}$$

Following (Ferland et al., 2006), the last equality of (A.4) follows from:

$$\begin{aligned} &\mathbb{E}(\alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j Z_{t-j}) = \mathbb{E}(\alpha_0 D^{-1}(1)) + \mathbb{E}(\sum_{j=1}^{\infty} \psi_j Z_{t-j}) \\ &= \alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j \mathbb{E}(Z_{t-j}) \\ &= \alpha_0 D^{-1}(1) + \mu H(1) = \alpha_0 D^{-1}(1) + \mu D^{-1}(1) G(1). \end{aligned} \tag{A.7}$$

From (A.5), a necessary condition for the second-order stationarity of $\{Z_t\}$ is: $\left(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j\right) > 0$. $\square$

A.2. Proof of the results in Section 3

Before proving Proposition 2, we introduce a suitable thinning operation and state a preliminary result. We denote with $\rho_{\theta,\lambda} \circ$ the quasi-binomial thinning operator, such that it follows a QB$(\rho, \theta/\lambda, x)$ distribution. See Alzaid and Al-Osh (1993) and Weiß (2008) and Supplementary Material for a definition.

Proposition A.1. *If* X *follows a* $GP(\lambda, \theta)$ *distribution and quasi-binomial thinning is performed independently on* X, *then* $\rho_{\theta,\lambda} \circ X$ *has a* $GP(\rho\lambda, \theta)$ *distribution.*

Proof. See Alzaid and Al-Osh (1993). $\square$

Proof of Proposition 2. It easy to show that, both $X_t^{(n)}$ and $Y_t^{(n)}$ in Eq. (7) and (8) admit the representation

$$X_t^{(n)} = (1-\lambda)U_{1t} + (1-\lambda)\sum_{i=1}^{n} \varphi_{1i}^{(t-i)} \circ \left(\frac{X_{t-i}^{(n-i)}}{1-\lambda}\right), \quad n > 0 \tag{A.8}$$

and

$$Y_t^{(n)} = (1-\lambda)U_{2t} + (1-\lambda)\sum_{i=1}^{n} \varphi_{2i}^{(t-i)} \circ \left(\frac{Y_{t-i}^{(n-i)}}{1-\lambda}\right), \quad n > 0, \tag{A.9}$$

where $\varphi \circ X$ is the quasi-binomial thinning operation. Let $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$ then

$$\begin{aligned} Z_t^{(n)} &= (1-\lambda)U_{1t} + (1-\lambda)\sum_{i=1}^{n} \varphi_{1i}^{(t-i)} \circ X_{t-i}^{(n-i)} - (1-\lambda)U_{2t} \\ &- (1-\lambda)\sum_{i=1}^{n} \varphi_{2i}^{(t-i)} \circ Y_{t-i}^{(n-i)} \\ &= (1-\lambda)(U_{1t} - U_{2t}) + (1-\lambda)\sum_{i=1}^{n} \left((\varphi_{1i}^{(t-i)} \circ X_{t-i}^{(n-i)}) \right. \\ &\left. - (\varphi_{2i}^{(t-i)} \circ Y_{t-i}^{(n-i)})\right) \\ &= (1-\lambda)U_t + (1-\lambda)\sum_{i=1}^{n} \varphi_i^{(t-i)} \diamond Z_{t-i}^{(n-i)}. \quad \square \end{aligned}$$

Before proving Proposition 3, we recall the definition of a probability-generating function (pgf) and prove a preliminary result. Let $\mathbf{W} = (W_1, \ldots, W_k)$ be a random vector and $\mathbf{w} = (w_1, \ldots, w_k)'$ a realization of $\mathbf{W}$. The pgf $g_{\mathbf{W}}(\mathbf{t})$ of $\mathbf{W}$ is defined as

$$g_{\mathbf{W}}(\mathbf{t}) = \mathbb{E}\left(\prod_{i=1}^{k} t_i^{W_i}\right) = \sum_{\mathbf{w}\in\mathbb{N}^k} \mathbb{P}(\{\mathbf{W}=\mathbf{w}\}) \prod_{i=1}^{k} t_i^{w_i}, \quad \text{(A.10)}$$

where $\mathbf{t} = (t_1, \ldots, t_k)' \in \mathbb{C}^k$. The pgf has the following properties.

Proposition A.2. *Let $\mathbf{Z}^{(n)}_{1\ldots k} = (Z^{(n)}_1, \ldots, Z^{(n)}_k)$ be a subsequence of $\{Z^{(n)}_t\}_{t\in\mathbb{Z}}$ where, without loss of generality, we choose the first k periods. Let $\mathbf{X}^{(n)}_{1\ldots k} = (X^{(n)}_1, \ldots, X^{(n)}_k)$ and $\mathbf{Y}^{(n)}_{1\ldots k} = (Y^{(n)}_1, \ldots, Y^{(n)}_k)$ be such that $\mathbf{Z}^{(n)}_{1\ldots k} = (\mathbf{X}^{(n)}_{1\ldots k} - \mathbf{Y}^{(n)}_{1\ldots k})'$ then*

$$g_{\mathbf{Z}_{1\ldots k}}(\mathbf{t}) = g_{\mathbf{X}_{1\ldots k}}(\mathbf{t}) g_{\mathbf{Y}_{1\ldots k}}(\mathbf{t}^{-1}). \quad \text{(A.11)}$$

Proof of Proposition A.2. The result follows from:

$$\begin{aligned} g_{\mathbf{Z}}(\mathbf{t}) &= \mathbb{E}\left(\prod_{i=1}^{k} t_i^{Z_i}\right) = \mathbb{E}\left(\prod_{i=1}^{k} t_i^{X_i}\right) \mathbb{E}\left(\prod_{i=1}^{k} \frac{1}{t_i^{Y_i}}\right) \\ &= g_{\mathbf{X}}(\mathbf{t}) g_{\mathbf{Y}}(\mathbf{t}^{-1}). \end{aligned} \quad \square \quad \text{(A.12)}$$

Proof of Proposition 3. (i) To prove almost sure convergence for $\{Z^{(n)}_t\}$, we show that the difference of two sequences $\{X^{(n)}_t\}$ and $\{Y^{(n)}_t\}$ that enjoy almost sure convergence will also enjoy almost sure convergence. We know that $Z^{(n)}_t = X^{(n)}_t - Y^{(n)}_t$, where $X^{(n)}_t$ and $Y^{(n)}_t$ are two sequences of GP random variables. From Zhu (2012) we have

$$X_n(\omega) \xrightarrow{\text{a.s.}} X(\omega) \Longrightarrow \mathbb{P}(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1,$$

$$Y_n(\omega) \xrightarrow{\text{a.s.}} Y(\omega) \Longrightarrow \mathbb{P}(\{\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}) = 1.$$

Let $A = \{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ and $B = \{\omega \in \Omega \times \Omega : \lim_{n\to\infty}(aX_n(\omega) + bY_n(\omega)) = aX(\omega) + bY(\omega)\}$. Then $\forall a, b \in \mathbb{R}$, we show almost sure convergence of the sum $(aX_n(\omega) + bY_n(\omega))$. Let

$$\begin{aligned} &\int_\Omega \mathbb{I}_B(\omega) d\mathbb{P}(\omega) = \int_{(\Omega\cap A)\cup(\Omega\cap A^C)} \mathbb{I}_B(\omega) d\mathbb{P}(\omega) \\ &= \int_\Omega \mathbb{I}_B(\omega)\mathbb{I}_A(\omega) d\mathbb{P}(\omega) + \int_\Omega \mathbb{I}_B(\omega)\mathbb{I}_{A^C}(\omega) d\mathbb{P}(\omega) \\ &= \int_\Omega \mathbb{I}_B(\omega)\mathbb{I}_A(\omega) d\mathbb{P}(\omega) + \int_\Omega \mathbb{I}_{A^C}(\omega) \\ &\left(\int_\Omega \mathbb{I}_{B\cap A^C}(\omega) d\mathbb{P}(\omega|\omega')\right) d\mathbb{P}(\omega') = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^C) \underbrace{\mathbb{P}(A^C)}_{=0} \\ &= \mathbb{P}(\{\omega : \lim_{n\to\infty}(aX_n(\omega) + bY_n(\omega))\}) \\ &= \mathbb{P}(\{\omega : a \lim_{n\to\infty} X_n(\omega) = aX(\omega) - bY(\omega) + bY(\omega)\}) = 1. \end{aligned} \quad \text{(A.13)}$$

Therefore, if $X^{(\omega)}_n \xrightarrow{\text{a.s.}} X(\omega)$ and $Y^{(\omega)}_n \xrightarrow{\text{a.s.}} Y(\omega)$ which implies $aX_n(\omega) + bY_n(\omega) \xrightarrow{\text{a.s.}} aX(\omega) + bY(\omega)$, $\forall a, b \in \mathbb{R}$. Hence, for $a = 1$ and $b = -1$, this is true for the difference $Z^{(n)}_t = X^{(n)}_t - Y^{(n)}_t$.

(ii) To prove the mean-square limit we use $Z^{(n)}_t = X^{(n)}_t - Y^{(n)}_t$ and the following lemma.

Lemma A.3. *If $X_n(\omega)$ and $Y_n(\omega)$ have a mean-square limit $X_n(\omega) \xrightarrow{L^2} X(\omega)$, $Y_n(\omega) \xrightarrow{L^2} Y(\omega)$, and for their sum: $aX_n(\omega) + bY_n(\omega) \xrightarrow{L^2} aX(\omega) + bY(\omega)$, $\forall a, b \in \mathbb{R}$.*

Hence, by setting $a = 1$ and $b = -1$ in Lemma A.3 we obtain $X_n(\omega) - Y_n(\omega) \xrightarrow{L^2} X(\omega) - Y(\omega)$ and that $Z^{(n)}_t$ converges to Z_t in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

(iii) In order to show stationarity for $\{Z^{(n)}_t\}$, we follow a procedure similar to the one in Ferland et al. (2006). Let k and h be two positive integers, then following (Brockwell, Davis, & Fienberg, 1991), to prove the strict stationarity of $(Z^{(n)}_t)_{t\in\mathbb{Z}}$, one needs to show that

$$\begin{aligned} &\mathbf{Z}^{(n)}_{1+h\ldots k+h} = (Z^{(n)}_{1+h}, \ldots, Z^{(n)}_{k+h})' \\ &\text{and } \mathbf{Z}^{(n)}_{1\ldots k} = (Z^{(n)}_1, \ldots, Z^{(n)}_k)' \end{aligned} \quad \text{(A.14)}$$

have the same joint distribution, where vectors in Eq. (A.14) are written as $\mathbf{Z}^{(n)}_{1+h\ldots k+h} = (\mathbf{X}^{(n)}_{1+h\ldots k+h} - \mathbf{Y}^{(n)}_{1+h\ldots k+h})' = ((X^{(n)}_{1+h} - Y^{(n)}_{1+h}), \ldots, (X^{(n)}_{k+h} - Y^{(n)}_{k+h}))'$ and $\mathbf{Z}^{(n)}_{1\ldots k} = (\mathbf{X}^{(n)}_{1\ldots k} - \mathbf{Y}^{(n)}_{1\ldots k})' = ((X^{(n)}_1 - Y^{(n)}_1), \ldots, (X^{(n)}_k - Y^{(n)}_k))'$. To show that both vectors have the same pgf, we first write the pgfs of X, Y, and Z, as shown above, and use the properties of the pgf given in Proposition A.2.

$$\begin{aligned} g_{\mathbf{X}^{(n)}_{1\ldots k}}(\mathbf{t}) &= \mathbb{E}\left(\prod_{j=1}^{k} t_j^{X^{(n)}_j}\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\mathbf{X}^{(n)}_{1\ldots k} | \mathbf{U}_{1,1-n\ldots k}\right)\left(\prod_{j=1}^{k} t_j^{X^{(n)}_j}\right)\right) \\ &= \sum_{\mathbf{v_1}\in\mathbb{N}^{(k+n)}} \mathbb{E}\left(\mathbf{X}^{(n)}_{1\ldots k} | \mathbf{U}_{1,1-n\ldots k} = \mathbf{v}_1\right) \\ &\left(\prod_{j=1}^{k} t_j^{X^{(n)}_j}\right) \mathbb{P}\left(\mathbf{U}_{1,1-n\ldots k} = \mathbf{v}_1\right) \end{aligned} \quad \text{(A.15)}$$

$$\begin{aligned} g_{\mathbf{Y}^{(n)}_{1\ldots k}}(\mathbf{t}) &= \sum_{\mathbf{v_2}\in\mathbb{N}^{(k+n)}} \mathbb{E}\left(\mathbf{Y}^{(n)}_{1\ldots k} | \mathbf{U}_{2,1-n\ldots k} = \mathbf{v}_2\right) \\ &\left(\prod_{j=1}^{k} t_j^{Y^{(n)}_j}\right) \mathbb{P}\left(\mathbf{U}_{2,1-n\ldots k} = \mathbf{v}_2\right) \end{aligned} \quad \text{(A.16)}$$

$$\begin{aligned} G_{\mathbf{Z}^{(n)}_{1\ldots k}}(\mathbf{t}) &= \sum_{\mathbf{v}\in\mathbb{N}^{(k+n)}} \mathbb{E}\left(\mathbf{Z}^{(n)}_{1\ldots k} | \mathbf{U}_{1-n\ldots k} = \mathbf{v}\right) \\ &\left(\prod_{j=1}^{k} t_j^{(X^{(n)}_j - Y^{(n)}_j)}\right) \mathbb{P}\left(\mathbf{U}_{1-n\ldots k} = \mathbf{v}\right). \end{aligned} \quad \text{(A.17)}$$

By the thinning representation, for any given value $\mathbf{u}_{1,t-n\ldots t+k} = (u_{1,t-n}, \ldots, u_{1,t+k})'$ of the vector $\mathbf{U}_{1,t-n\ldots t+k} = (U_{1,t-n}, \ldots, U_{1,t+k})'$ and $\mathbf{u}_{2,t-n\ldots t+k} = (u_{2,t-n}, \ldots, u_{2,t+k})'$ of the vector $\mathbf{U}_{2,t-n\ldots t+k} = (U_{2,t-n}, \ldots, U_{2,t+k})'$, the components of the vectors $(X^{(n)}_1, \ldots, X^{(n)}_k)'$ and $(Y^{(n)}_1, \ldots, Y^{(n)}_k)'$ are computed using a set of well-determined variables from the sequences $\mathcal{V}_{1,\tau,\eta}$ and $\mathcal{V}_{2,\tau,\eta}$,

where $\tau = t-n, \dots, t+k-1$ and $\eta = 1, \dots, n$. Therefore, if $\mathbf{U}_{1,t-n\dots t+k}$ and $\mathbf{U}_{1,t-n+h\dots t+k+h}$ are both fixed to the same value $\mathbf{v_1}$ and $\mathbf{U}_{2,t-n\dots t+k}$ and $\mathbf{U}_{2,t-n+h\dots t+k+h}$ are both fixed to the same value $\mathbf{v}_2$, it follows that the conditional distribution of $\mathbf{Z}^{(n)}_{1+h\dots k+h} = ((X^{(n)}_{1+h} - Y^{(n)}_{1+h}), \dots, (X^{(n)}_{k+h} - Y^{(n)}_{k+h}))'$ and $\mathbf{Z}^{(n)}_{1\dots k} = ((X^{(n)}_1 - Y^{(n)}_1), \dots, (X^{(n)}_k - Y^{(n)}_k))'$ given $\mathbf{U}_{t-n\dots t+k}$ and $\mathbf{U}_{t-n+h\dots t+k+h}$, are the same. Accordingly,

$$\mathbb{E}\left(\mathbf{Z}^{(n)}_{1+h\dots k+h}|\mathbf{U}_{1-n+h\dots k+h} = \mathbf{v}\right)\left(\prod_{j=1}^{k} t_j^{Z^{(n)}_{j+h}}\right)$$
$$= \mathbb{E}\left(\mathbf{Z}^{(n)}_{1\dots k}|\mathbf{U}_{1-n\dots k} = \mathbf{v}\right)\left(\prod_{j=1}^{k} t_j^{Z^{(n)}_{j}}\right)$$

and, since $\mathbb{P}(\{\mathbf{U}_{1-n+h\dots k+h} = \mathbf{v}\}) = \mathbb{P}(\{\mathbf{U}_{1-n\dots k} = \mathbf{v}\})$, it is possible to write

$$\begin{aligned} g_{\mathbf{Z}^{(n)}_{1\dots k}}(\mathbf{t}) &= \sum_{\mathbf{v}\in\mathbb{Z}^{(k+n)}} \mathbb{E}\left(\mathbf{Z}^{(n)}_{1+h\dots k+h}|\mathbf{U}_{1-n+h\dots k+h} = \mathbf{v}\right) \\ &\quad \left(\prod_{j=1}^{k} t_j^{Z^{(n)}_{j+h}}\right)\mathbb{P}(\{\mathbf{U}_{1-n+h\dots k+h} = \mathbf{v}\}) \\ &= g_{\mathbf{Z}^{(n)}_{1+h\dots k+h}}(\mathbf{t}) \end{aligned}$$

and conclude that $\mathbf{Z}^{(n)}_{1+h\dots k+h}$ and $\mathbf{Z}^{(n)}_{1\dots k}$ have the same joint distribution. □

Proof of Proposition 4. The process $(Z_t)_{t\in\mathbb{Z}}$ is stationary since it is the a.s. limit of the stationary process $Z^{(n)}_t$. We now prove the first- and second-order moments are finite. Define $Z^{(n)}_t = X^{(n)}_t - Y^{(n)}_t$, where $X^{(n)}_t$ and $Y^{(n)}_t$ are finite sums of independent Generalized Poisson variables. Then it follows that $Z^{(n)}_t$ is a finite sum of Generalized Poisson difference variables. As shown by Zhu (2012), the first two moments of X_t and Y_t are finite: $\mathbb{E}[X_t] = \mu_X \le C_1$, $\mathbb{E}[Y_t] = \mu_Y \le C_1'$, $\mathbb{V}[X_t] = \sigma^2_X \le C_2$, $\mathbb{V}[Y_t] = \sigma^2_Y \le C_2'$, therefore, $\mathbb{E}[Z_t] = \mathbb{E}[X_t] - \mathbb{E}[Y_t] = \mu_X - \mu_Y \le \mu_X + \mu_Y \le C_1 + C_1'$ is finite and $\mathbb{V}[X_t - Y_t] = \mathbb{V}[X_t] + \mathbb{V}[X_t] = \sigma^2_X + \sigma^2_Y \le C_2 + C_2'$ is also finite, where $\mathbb{C}ov(X_t, Y_t) = 0$ since X_t and Y_t are independent and where C_i and C_i', $i = 1, 2$ are constants. □

Proof of Proposition 5. To verify that the distributional properties of the sequence are satisfied, we will follow the same arguments as in Ferland et al. (2006) but adjust for our sequence. Given $\mathcal{F}_{t-1} = \sigma(\{Z_u\}_{u\le t-1})$, for $t \in \mathbb{Z}$, let

$$\mu_t = \alpha_0 D^{-1}(1) + \sum_{j=1}^{n} \psi_j Z_{t-j}.$$

The sequence $\{\mu_t\}$ satisfies

$$\mu_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i} + \sum_{j=1}^{q} b_j \mu_{t-j}. \tag{A.18}$$

Moreover, recalling that $Z_t = X_t - Y_t$, for a fixed t, we can consider three sequences, $\{r^{(n)}_{1t}\}_{n\in\mathbb{N}}$, $\{r^{(n)}_{2t}\}_{n\in\mathbb{N}}$ and $\{r^{(n)}_t\}_{n\in\mathbb{N}}$, defined by

$$r^{(n)}_{1t} = (1-\lambda)U_{1t} + (1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{X_{t-i}} V_{1t-i,i,k} \tag{A.19}$$

$$r^{(n)}_{2t} = (1-\lambda)U_{2t} + (1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{Y_{t-i}} V_{2t-i,i,k}. \tag{A.20}$$

and

$$r^{(n)}_t = r^{(n)}_{1t} - r^{(n)}_{2t}. \tag{A.21}$$

As shown by Ferland et al. (2006), there is a subsequence $\{n_k\}$ such that $r^{(n_k)}_t$ converges almost surely to Z_t. We know that

$$X_t - r^{(n)}_{1t} = (X_t - X^{(n)}_t) + (X^{(n)}_t - r^{(n)}_{1t}) \tag{A.22}$$

and

$$Y_t - r^{(n)}_{2t} = (Y_t - Y^{(n)}_t) + (Y^{(n)}_t - r^{(n)}_{2t}). \tag{A.23}$$

Since $X^{(n)}_t \xrightarrow{a.s.} X_t$ and $Y^{(n)}_t \xrightarrow{a.s.} Y_t$, we know that the first term in both Eq. (A.22) and (A.23) goes to zero. Therefore, we can write

$$\begin{aligned} Z_t - r^{(n)}_t &= (X_t - Y_t) - (r^{(n)}_{1t} - r^{(n)}_{2t}) \\ &= \left[(X_t - X^{(n)}_t) - (Y_t - Y^{(n)}_t)\right] \\ &\quad + \left[(X^{(n)}_t - r^{(n)}_{1t}) - (Y^{(n)}_t - r^{(n)}_{2t})\right] \\ &= (Z_t - Z^{(n)}_t) + \left[(X^{(n)}_t - Y^{(n)}_t) - (r^{(n)}_{1t} - r^{(n)}_{2t})\right] \\ &= (Z_t - Z^{(n)}_t) + \left[Z^{(n)}_t - (r^{(n)}_{1t} - r^{(n)}_{2t})\right], \end{aligned} \tag{A.24}$$

and, as before, $(Z_t - Z^{(n)}_t)$ goes to zero since we have proven almost sure convergence. We have now to show that the second term in the last line of Eq. (A.24) goes to zero. For this purpose, we need to find a sequence $W^{(n)}_t = (r^{(n)}_{1t} - r^{(n)}_{2t}) - Z^{(n)}_t$ that converges almost surely to zero. For this reason, it is more suitable to rewrite the previous sequence as $W^{(n)}_t = (r^{(n)}_{1t} - r^{(n)}_{2t}) - (X_t - Y_t) = (r^{(n)}_{1t} - X_t) - (r^{(n)}_{2t} - Y_t)$. Ferland et al. (2006) show that

$$\lim_{n\to\infty} \mathbb{E}\left((r^{(n)}_{1t} - X_t)\right) = 0, \quad \lim_{n\to\infty} \mathbb{E}\left((r^{(n)}_{2t} - Y_t)\right) = 0 \tag{A.25}$$

therefore, we can conclude that

$$\lim_{n\to\infty} \mathbb{E}\left((r^{(n)}_t - Z_t)\right) = 0. \tag{A.26}$$

Eq. (A.26) implies that $W^{(n)}_t$ converges to zero in L^1, therefore there exists a sub-sequence $W^{(n_k)}_t$ converging almost surely to the same limit. From this, the distributional properties of X_t are satisfied. Since $r^{(n_k)}_{1t} \xrightarrow{a.s.} X_t$ and $r^{(n_k)}_{2t} \xrightarrow{a.s.} Y_t$, it is also true that $r^{(n_k)}_t \xrightarrow{a.s.} Z_t$. Hence,

$$r^{(n)}_t|\mathcal{F}_{t-1} \xrightarrow{a.s.} Z_t|\mathcal{F}_{t-1}.$$

However, $r^{(n)}_t|\mathcal{F}_{t-1} = (r^{(n)}_{1t} - r^{(n)}_{2t})|\mathcal{F}_{t-1}$ and from Zhu (2012) we know that both $r^{(n)}_{1t}$ and $r^{(n)}_{2t}$ have a Generalized

Poisson distribution. Since the difference between two GP random variables is GPD distributed, we can write

$$r_t^{(n)}|\mathcal{F}_{t-1} \sim GPD\left(\alpha_0 D^{-1}(1) + \sum_{j=1}^{n} \psi_j Z_{t-j}\right) \tag{A.27}$$

and conclude that $Z_t|\mathcal{F}_{t-1} \sim GPD(\tilde{\mu}_t, \tilde{\sigma}_t^2, \lambda)$. □

Proof of the results in Example 1. For $k \geq 2$, $\gamma_Z(k) = \alpha_1\gamma_Z(k-1) = \alpha_1^{k-1}\gamma_Z(1)$, and for $k = 1$, $\gamma_Z(1) = \mathbb{C}ov(Z_t, Z_{t-1}) = \alpha_1\gamma_Z(0) = \alpha_1\mathbb{V}(Z_t) = \alpha_1(\phi^3\mathbb{E}(\sigma_t^2)) + \alpha_1\mathbb{V}(\mu_t)$. For $k \geq 1$ we have $\gamma_\mu(k) = \alpha_1\gamma_\mu(k-1) = \alpha_1^k\mathbb{V}(\mu_t)$. For $k = 0$, $\gamma_\mu(k) = \mathbb{V}(\mu_t) = \alpha_1\gamma_Z(1) = \alpha_1\left\{\alpha_1(\phi^3\mathbb{E}(\sigma_t^2) + \mathbb{V}(\mu_t))\right\} = \alpha_1^2(\phi^3\mathbb{E}(\sigma_t^2)) + \alpha_1^2\mathbb{V}(\mu_t)$. Therefore,

$$\mathbb{V}(\mu_t) = \frac{\alpha_1^2(\phi^3\mathbb{E}(\sigma_t^2))}{1-\alpha_1^2} \tag{A.28}$$

$$\begin{aligned}\mathbb{V}(Z_t) &= \phi^3\mathbb{E}(\sigma_t^2) + \mathbb{V}(\mu_t)\\ &= \phi^3\mathbb{E}(\sigma_t^2) + \frac{\alpha_1^2[\phi^3\mathbb{E}(\sigma_t^2)]}{1-\alpha_1^2} = \frac{\phi^3\mathbb{E}(\sigma_t^2)}{1-\alpha_1^2},\end{aligned} \tag{A.29}$$

where $\phi = 1/(1-\lambda)$. Finally, the autocorrelations are derived as follows:

$$\rho_\mu(k) = \frac{\gamma_\mu(k)}{\mathbb{V}(\mu_t)} = \frac{\alpha_1^k\,\mathbb{V}(\mu_t)}{\mathbb{V}(\mu_t)} = \alpha_1^k \tag{A.30}$$

$$\begin{aligned}\rho_Z(k) &= \frac{\gamma_Z(k)}{\mathbb{V}(Z_t)} = \alpha_1^{k-1}\,\gamma_Z(1)\frac{1-\alpha_1^2}{\phi^3\mathbb{E}(\sigma_t^2)}\\ &= \alpha_1^{k-1}\,\frac{\alpha_1(1-\alpha_1^2)\phi^3\mathbb{E}(\sigma_t^2) + \alpha_1^3\phi^3\mathbb{E}(\sigma_t^2)}{1-\alpha_1^2}\,\frac{1-\alpha_1^2}{\phi^3\mathbb{E}(\sigma_t^2)}\\ &= \alpha_1^k. \quad \square\end{aligned} \tag{A.31}$$

Proof of the results in Example 2. For $k \geq 2$, $\gamma_Z(k) = \alpha_1\gamma_Z(k-1) + \beta_1\gamma_Z(k-1) = (\alpha_1+\beta_1)^{k-1}\gamma_Z(1)$. For $k = 1$, $\gamma_Z(1) = \mathbb{C}ov(Z_t, Z_{t-1}) = \alpha_1\gamma_Z(0) + \beta_1\gamma_\mu(0) = \alpha_1\mathbb{V}(Z_t) + \beta_1\mathbb{V}(\mu_t) = \alpha_1[\phi^3\mathbb{E}(\sigma_t^2)] + (\alpha_1+\beta_1)\mathbb{V}(\mu_t)$. For $k \geq 1$ we have $\gamma_\mu(k) = \alpha_1\gamma_\mu(k-1) + \beta_1\gamma_\mu(k-1) = (\alpha_1+\beta_1)^k\mathbb{V}(\mu_t)$. For $k = 0$, $\gamma_\mu(k) = \mathbb{V}(\mu_t) = \alpha_1\gamma_Z(1) + \beta_1\gamma_\mu(1) = \alpha_1\left\{\alpha_1(\phi^3\mathbb{E}(\sigma_t^2)) + (\alpha_1+\beta_1)\mathbb{V}(\mu_t)\right\} + \beta_1\left((\alpha_1+\beta_1)\mathbb{V}(\mu_t)\right) = \alpha_1^2(\phi^3\mathbb{E}(\sigma_t^2)) + (\alpha_1+\beta_1)^2\mathbb{V}(\mu_t)$. We conclude that:

$$\begin{aligned}\mathbb{V}(\mu_t) &= \frac{\alpha_1^2(\phi^3\mathbb{E}(\sigma_t^2))}{1-(\alpha_1+\beta_1)^2},\\ \mathbb{V}(Z_t) &= \phi^3\mathbb{E}(\sigma_t^2) + \mathbb{V}(\mu_t) = \frac{\phi^3\mathbb{E}(\sigma_t^2)(1-(\alpha_1+\beta_1)^2+\alpha_1^2)}{1-(\alpha_1+\beta_1)^2}.\end{aligned} \tag{A.32}$$

The autocorrelations are derived as follows:

$$\rho_\mu(k) = \frac{\gamma_\mu(k)}{\mathbb{V}(\mu_t)} = \frac{(\alpha_1+\beta_1)^k\,\mathbb{V}(\mu_t)}{\mathbb{V}(\mu_t)} = (\alpha_1+\beta_1)^k \tag{A.33}$$

$$\begin{aligned}\rho_Z(k) &= \frac{\gamma_Z(k)}{\mathbb{V}(Z_t)}\\ &= (\alpha_1+\beta_1)^{k-1}\,\gamma_Z(1)\frac{1-(\alpha_1+\beta_1)^2}{\phi^3\mathbb{E}(\sigma_t^{2^*})(1-(\alpha_1+\beta_1)^2+\alpha_1^2)}\\ &= (\alpha_1+\beta_1)^{k-1}\,\frac{\alpha_1(1-\beta_1(\alpha_1+\beta_1))}{1-(\alpha_1+\beta_1)^2+\alpha_1^2} \quad \square\end{aligned} \tag{A.34}$$

Table B.1
Value of Geweke's test statistics for the three parameters of the GPD-INGARCH(1,1) and different choices of the test function $g(x)$.

$g(x)$	α	β	λ
x	−1.32	0.14	1.17
x^2	0.0881	0.3472	1.5581
x^3	1.4672	0.5406	1.8321

Appendix B. Further details for the numerical illustration

B.1. Correct implementation test

Fig. B.1 shows Geweke's statistics evaluated on the first n MCMC samples, $n = 1, \ldots, 2000$, for the parameters α, β and λ and the first three moments. A graphical inspection indicates convergence of the statistics. Table B.1 reports the statistics on the 2,000 samples for the different parameters (columns) and choices of the test function (rows). The absolute value of the statistics Z is always below 2.58, the critical value of Geweke's statistics at the 1% level. Thus, the null hypothesis of the correct implementation of the MCMC is accepted. See Section S.2 in Supplement Material for background material on Geweke's statistics.

B.2. Sampling efficiency

We consider 400 samples from two GPD-INGARCH(1,1) simulation settings: one with low persistence and the other with high persistence. The first setting has parameters $\lambda = 0.4$, $\alpha_1 = 0.25$, $\beta_1 = 0.23$, $\alpha_0 = -0.2$ and $\phi = 22.78$, while the second setting with parameters $\lambda = 0.6$, $\alpha_1 = 0.53$, $\beta_1 = 0.25$, $\alpha_0 = -0.2$ and $\phi = 26.25$. We run the Gibbs sampler for 1,010,000 iterations, discard the first 10,000 draws to avoid dependence from initial conditions, and apply a thinning procedure with a factor of 100 to reduce the dependence between consecutive draws. The following figures show the posterior approximation of α_1, β_1 and λ. For illustrative purposes, we report in Figs. B.2 to B.4 the MCMC output for one MCMC draw before removing the burn-in sample and thinning. In contrast, in Figs. B.5 to B.7, we display the MCMC output after removing the burn-in sample and implementing thinning.

B.3. Prior sensitivity

This section compares inference results under two prior assumptions. For the diffuse prior, we assume the Dirichlet prior given in Eq. with hyper-parameter values $c_0 = 3$, $c_1 = 4$ and $c_2 = 3$. For the strongly informative prior, we assume a Dirichlet prior with $c_0 = 0.8$, $c_1 = 0.8$, and $c_2 = 0.8$ which assigns a large probability mass to values of $\alpha + \beta$ close to 1. The sensitivity to the prior choice has been assessed in the low and high persistence settings.

Fig. B.8 shows the contour lines of the prior and the MCMC samples from the posterior of α and β. Fig. B.9

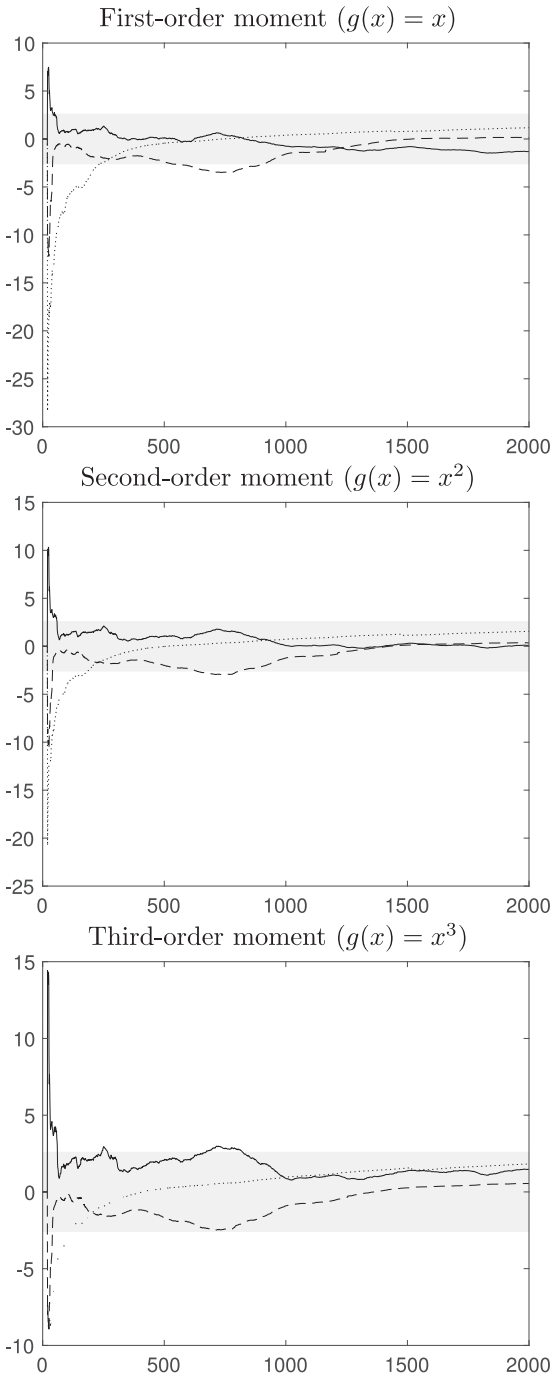

Fig. B.1. Plot of Geweke's statistics for correct implementation of the MCMC. Statistics are evaluated on the first n MCMC samples, $n = 1, \ldots, 2000$, for the parameters α (solid line), β (dashed line), and λ (dotted line) and the first three moments (different plots). Grey is the acceptance region of the test at the 1% level.

shows the mean and the 2.5% and 97.5% quantiles of the posterior distribution of $\alpha + \beta$ approximated with an increasing number of MCMC samples (after thinning and burn-in) (see Fig. C.3).

In all scenarios, the information from the data dominates the posterior, which is not very sensitive to the prior choice.

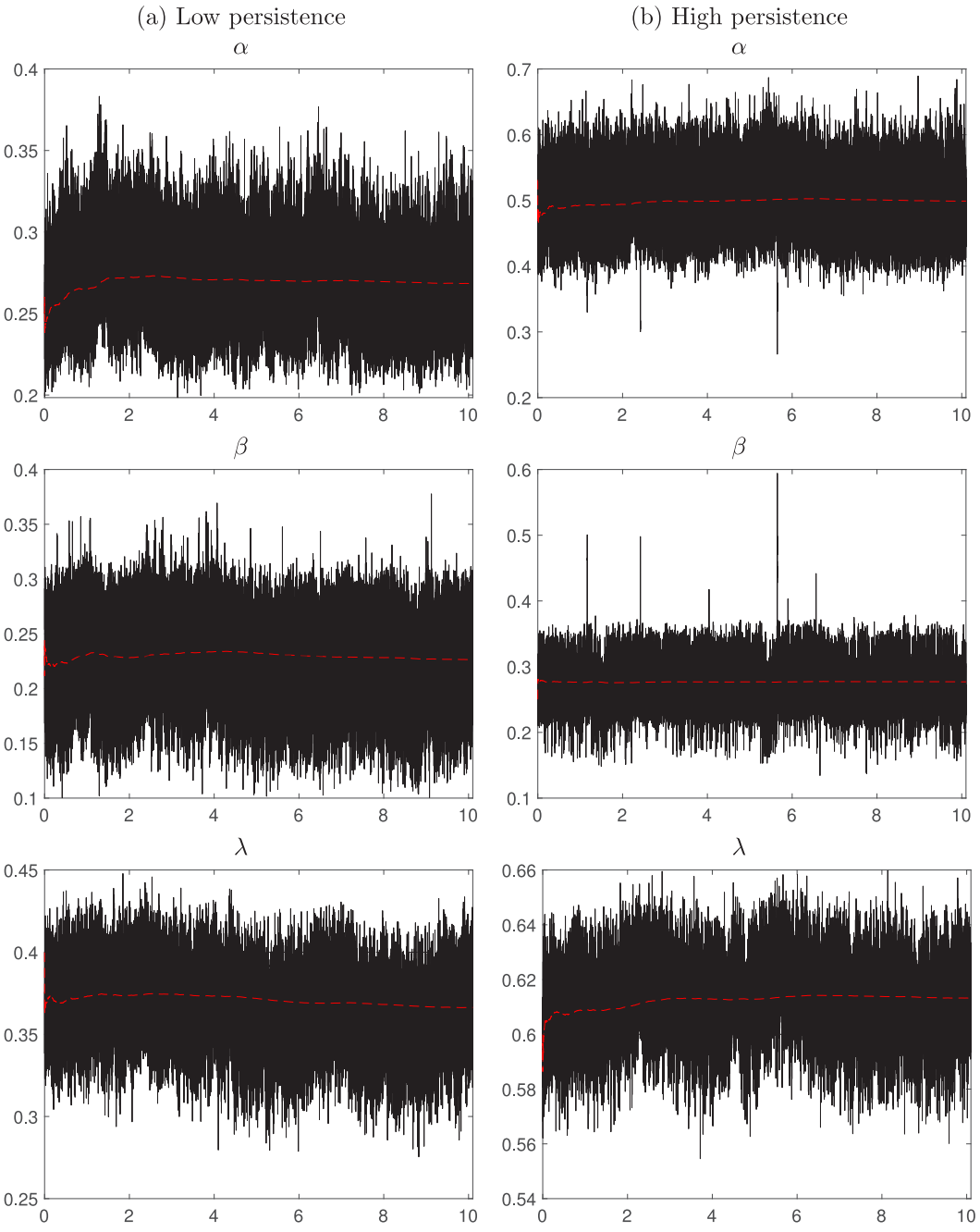

Fig. B.2. MCMC trace plots for the parameters in both settings: low and high persistence.

Appendix C. Further details for the real-data applications

C.1. Preliminary analysis

In this section, the descriptive statistics and the results of the stationarity and normality tests are reported for Schiphol's (Table C.1) and Cyber threats (Table C.2) datasets (see Fig. C.4).

C.2. Posterior inference

The following reports the posterior histograms and the 1-step-ahead forecasts for Schiphol's and Cyber threats datasets. The mean square errors at different horizons, $h = 1, \ldots, 7$, are provided for the Generalized Poisson difference specifications, GPD-INGARCH and GPD-INARCH, and the standard Poisson difference specifications, PD-INGARCH and PD-INARCH (see Fig. C.6).

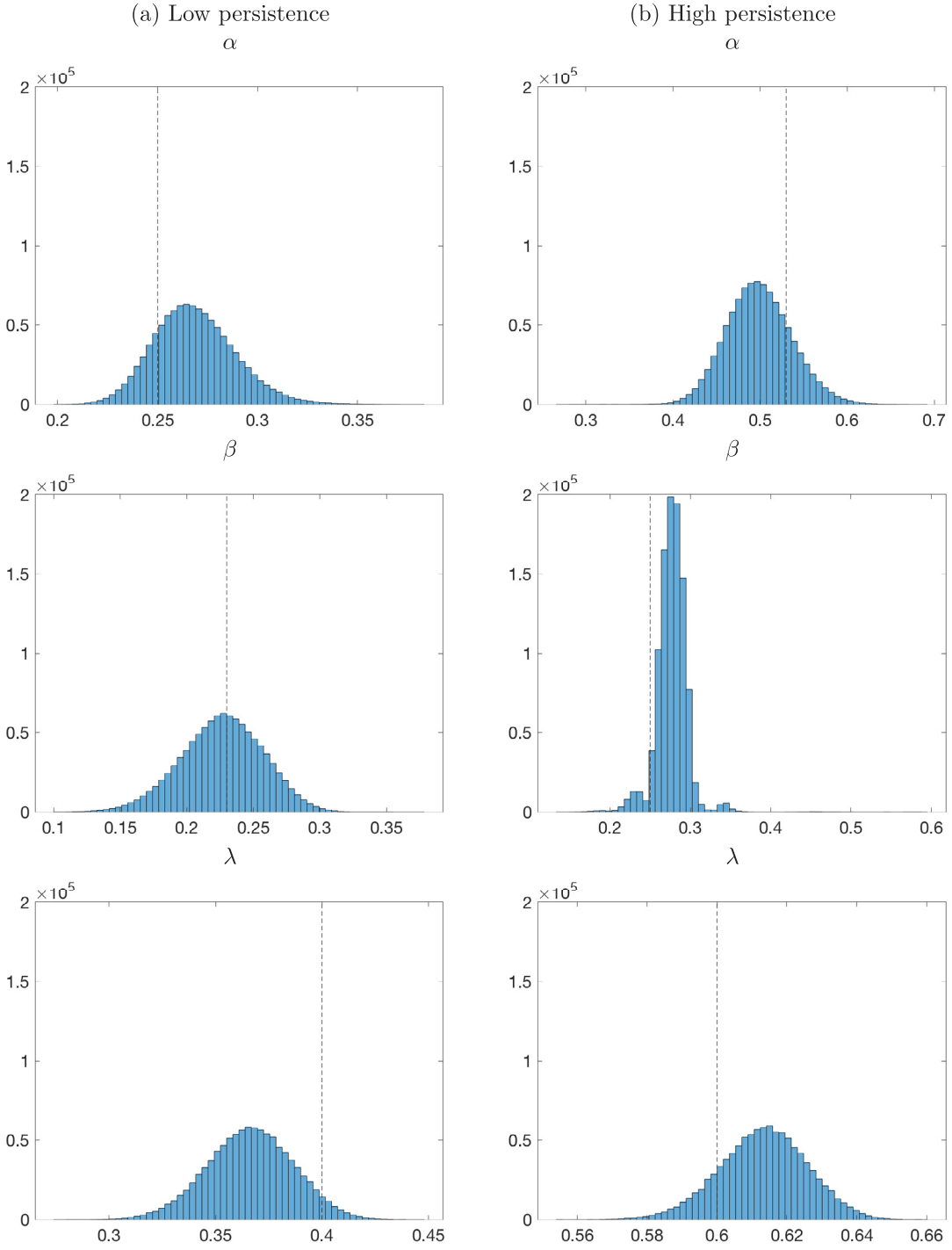

Fig. B.3. Histograms of MCMC draw for the parameters in both settings: low and high persistence.

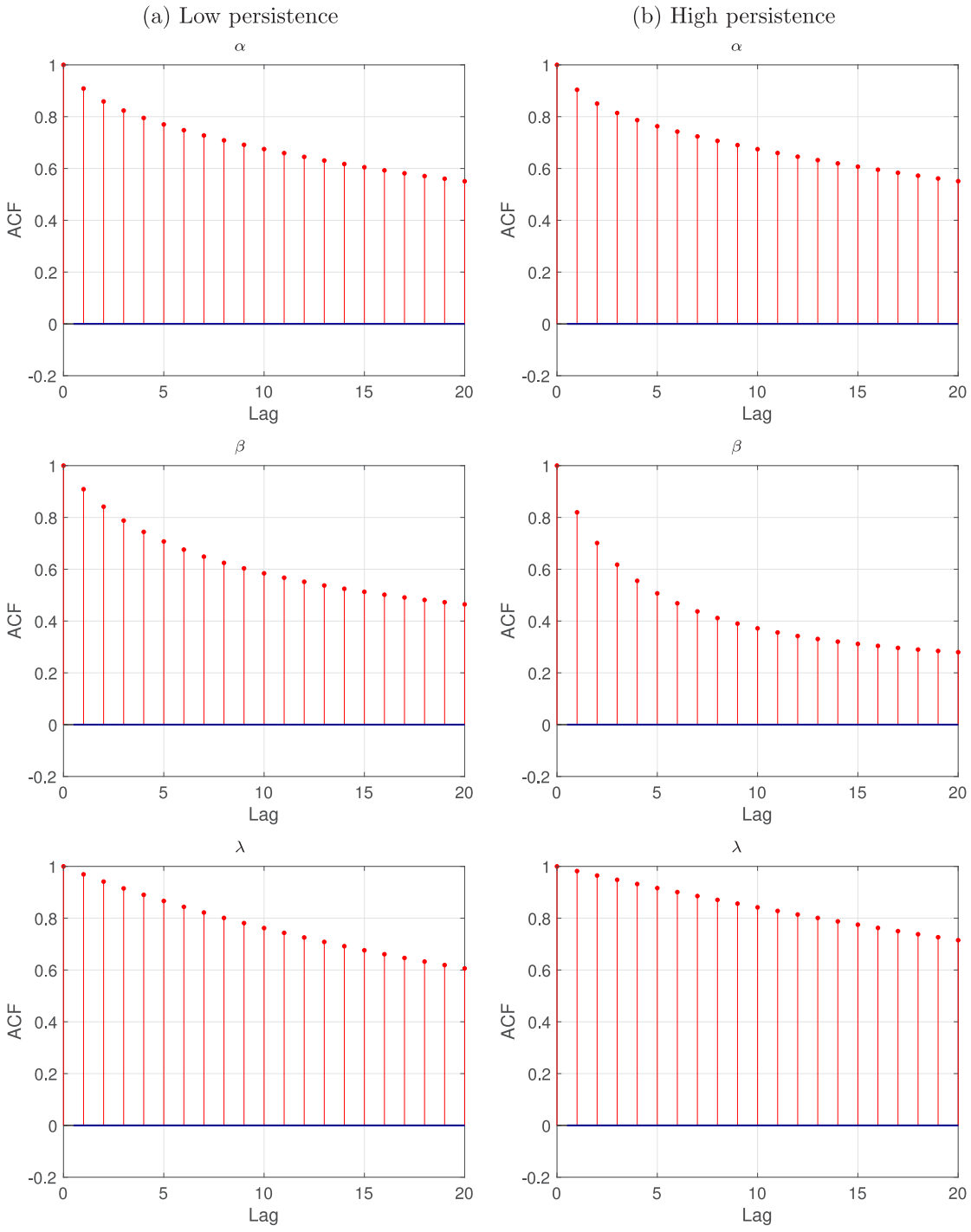

Fig. B.4. Autocorrelation functions for the parameters in both low and high persistence settings.

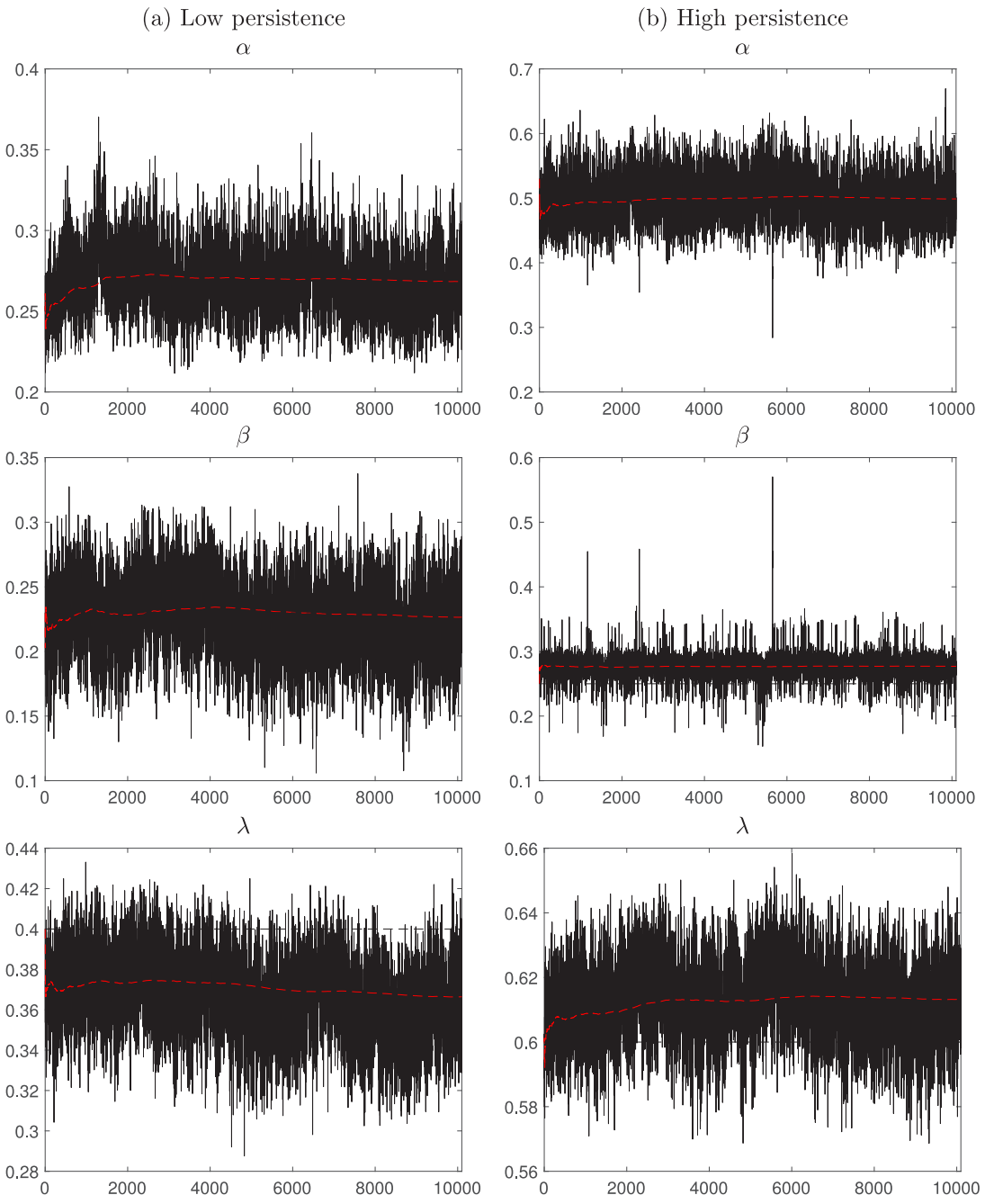

Fig. B.5. MCMC trace plots for the parameters in both settings: low and high persistence, after removing the burn-in sample and implementing thinning.

Table C.1
Descriptive statistics, stationary and normality tests for the car accident dataset. ADF is the Augmented Dickey–Fuller test with stationarity as the null hypothesis, PP is the Phillips–Perron test with stationarity as the null hypothesis, and KS is the Kolmogorov–Smirnov test where the null hypothesis states that the data comes from a Normal distribution.

	Mean	Variance	Skewness	Kurtosis	ADF	PP	KS
Car Accident	8.78	24.55	0.91	3.89	−9.21**	−15.66**	0.94**
Δ Car accident	−0.0055	39.71	0.06	3.59	−18.35**	−34.84**	0.36**

** The symbol means that the null hypothesis is rejected at 5% significance level.

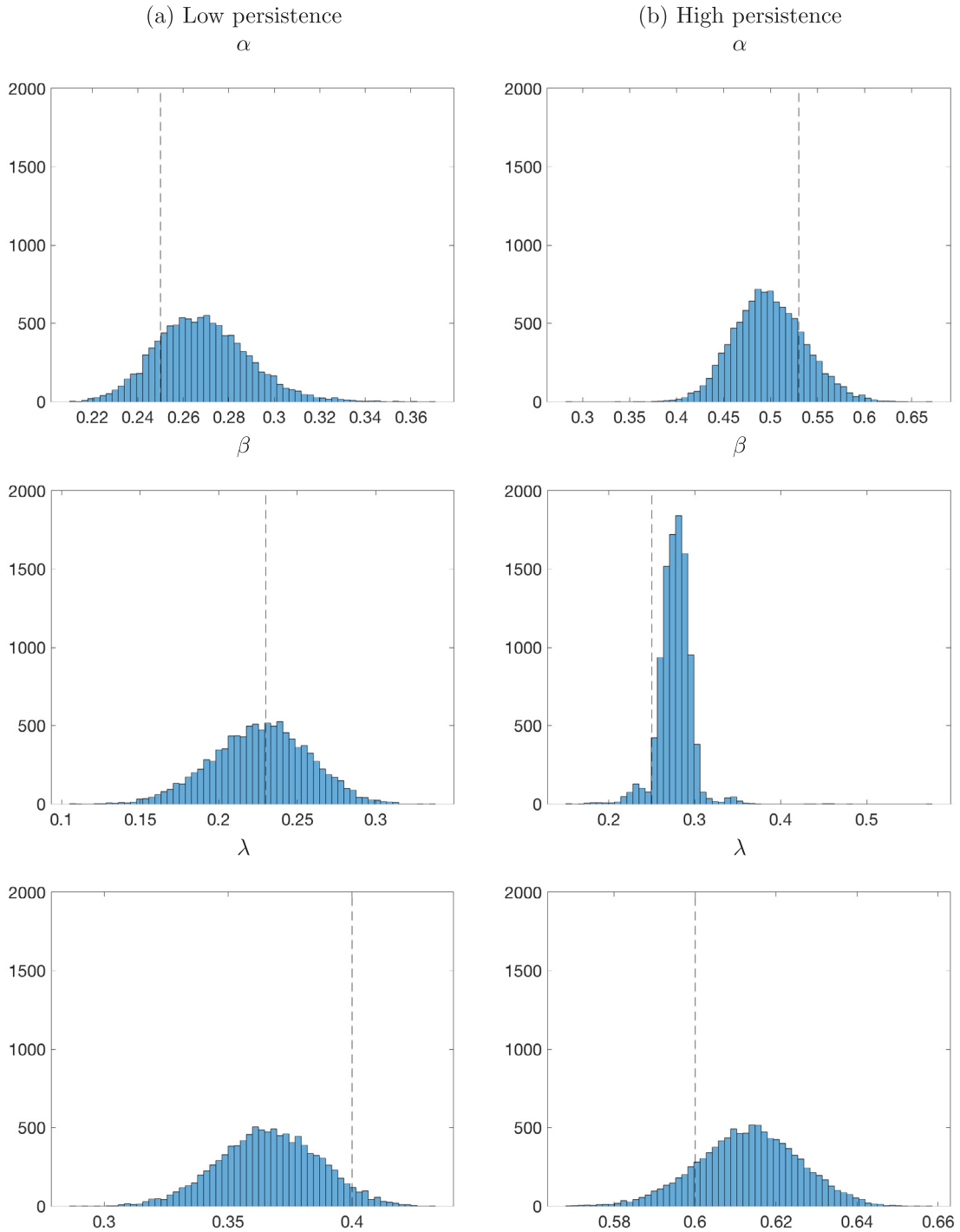

Fig. B.6. Histograms of MCMC draws for the parameters in both settings: low and high persistence, after removing the burn-in sample and implementing thinning.

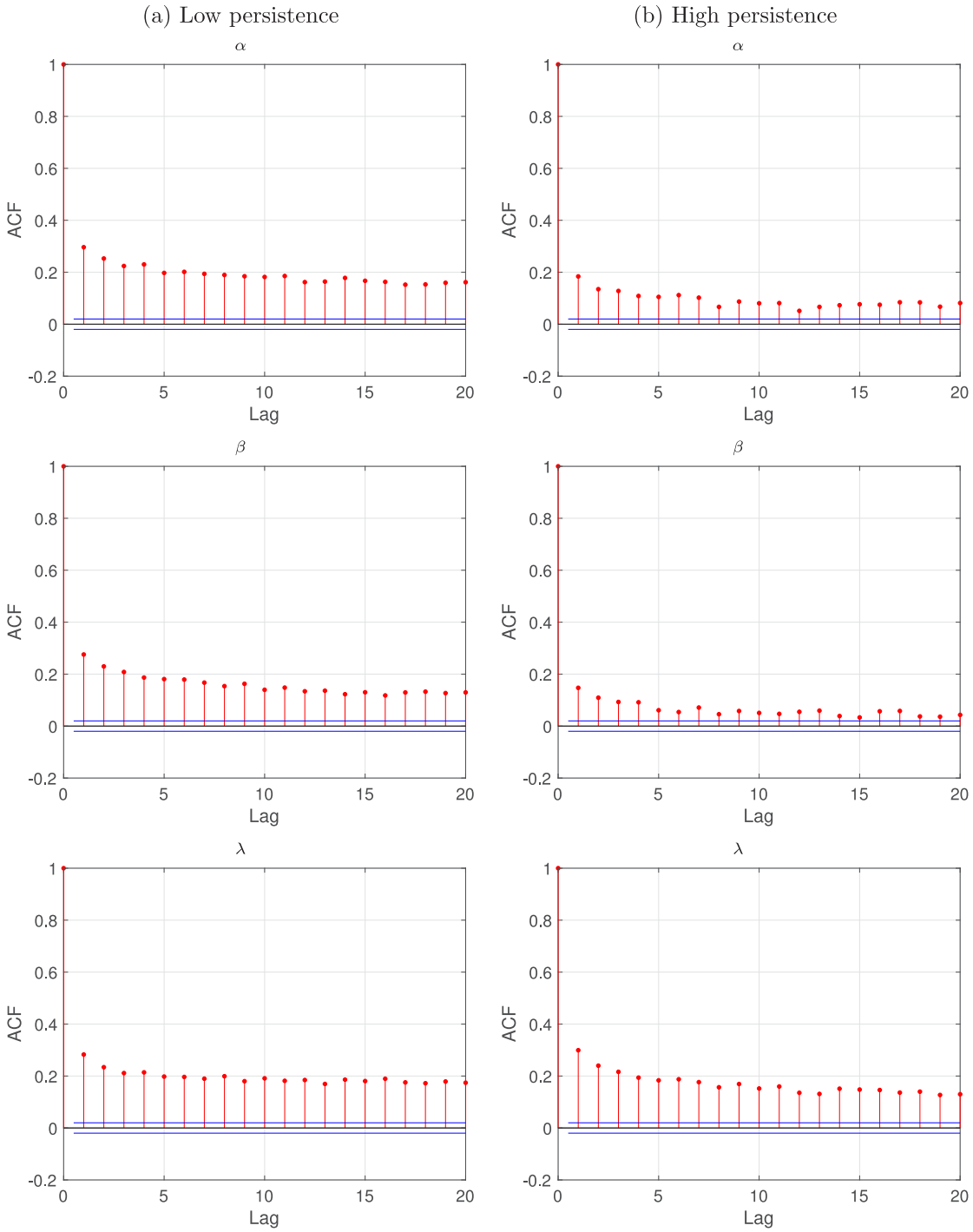

Fig. B.7. Autocorrelation functions for the parameters in low and high persistence settings, after removing the burn-in sample and implementing thinning.

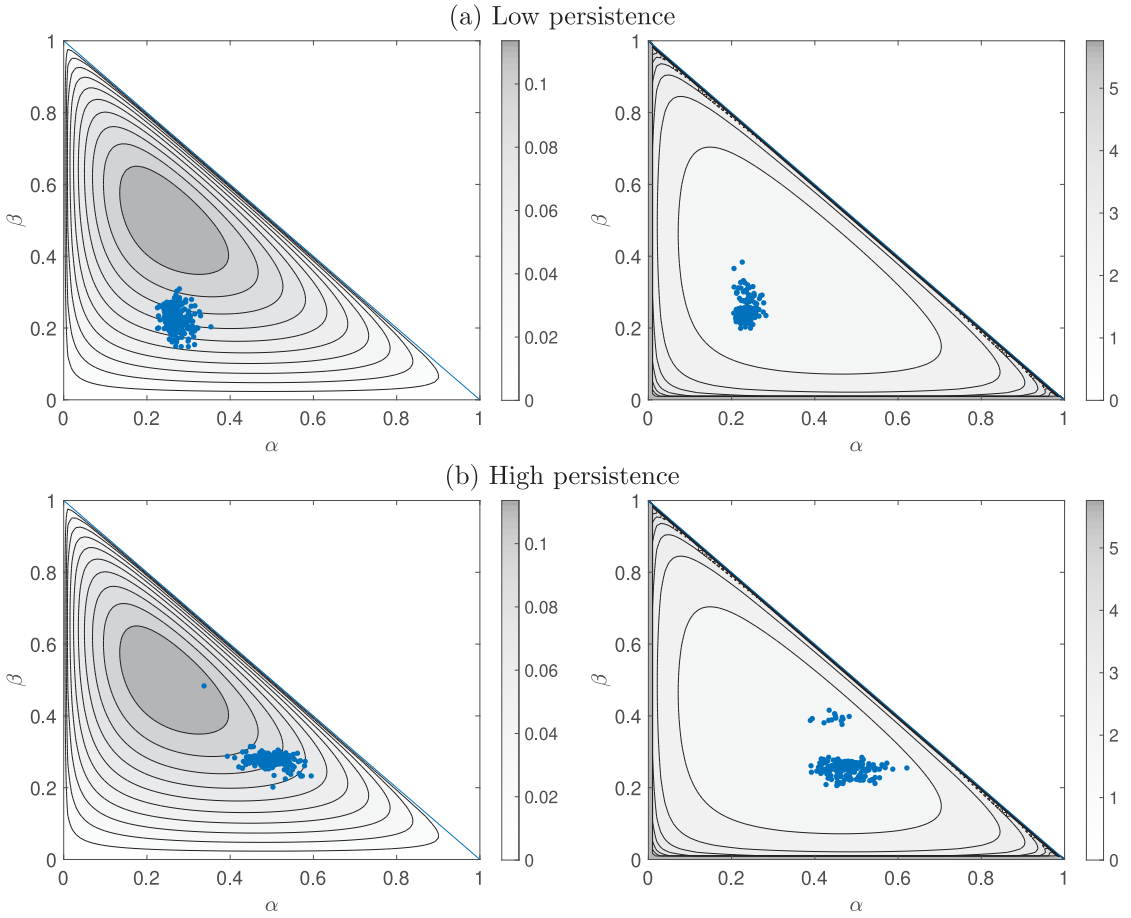

Fig. B.8. Log-density function (contour lines) and MCMC samples (blue dots) of α and β for the diffuse prior (left) and the strongly informative prior (right) in the low and high persistence settings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

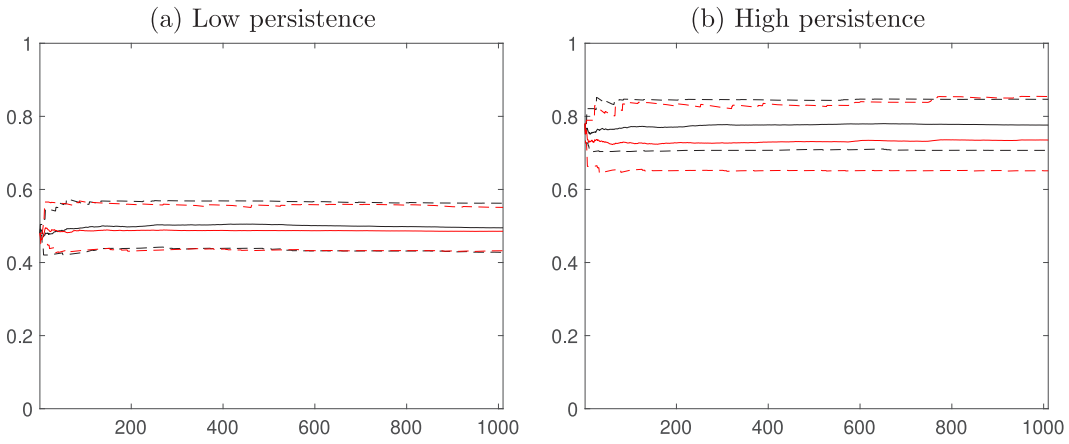

Fig. B.9. Posterior mean (solid) and quantiles (dashed) of $\alpha + \beta$ for an increasing number of MCMC iterations (horizontal axis) under a diffuse prior (black) and a strongly informative prior (red) assumptions in the low and high persistence settings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table C.2
Descriptive statistics, stationary and normality tests for the cyber dataset. ADF is the Augmented Dickey–Fuller test with stationarity as the null hypothesis, PP is the Phillips–Perron test with stationarity as the null hypothesis, and KS is the Kolmogorov–Smirnov test where the null hypothesis states that the data comes from a Normal distribution.

	Mean	Variance	Skewness	Kurtosis	ADF	PP	KS
Cyber	3.72	8.08	1.07	5.07	−23.16**	−26.98**	0.74**
Δ Cyber	9.14e−04	11.72	0.12	4.53	−30.89**	−48.91**	0.29**

** The symbol means that the null hypothesis is rejected at 5% significance level.

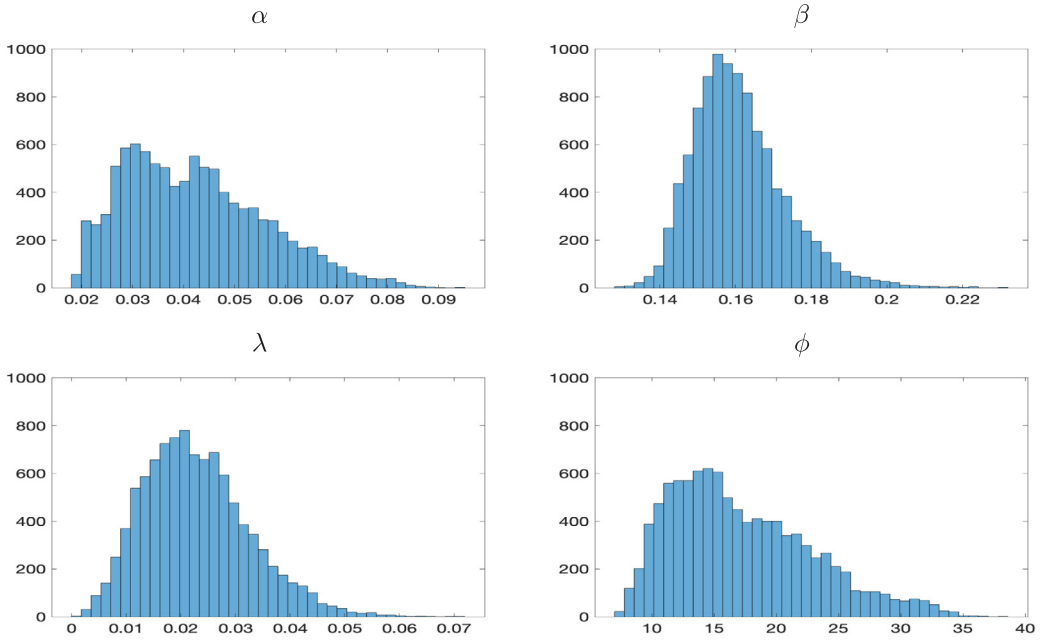

Fig. C.1. Histograms of MCMC draws for the parameters of Schiphol's accident data of Fig. 6.

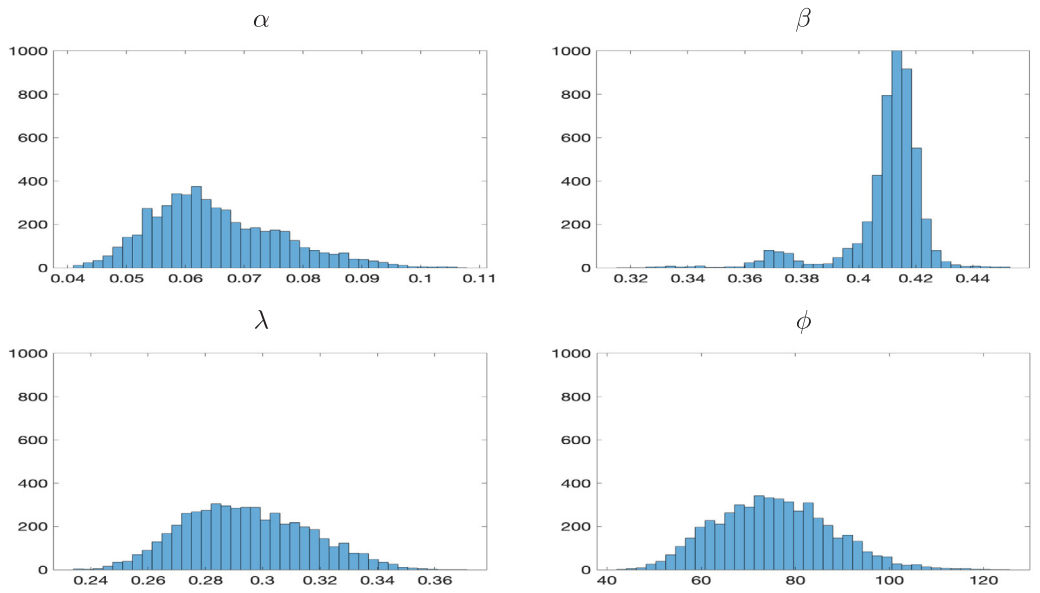

Fig. C.2. Histograms of MCMC draws for the parameters of the Cyber treats data of Fig. 8.

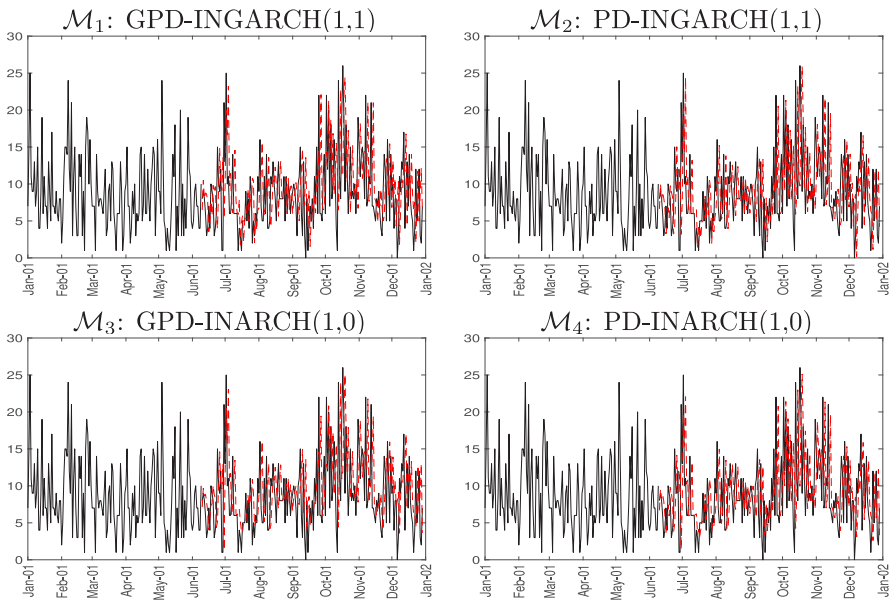

Fig. C.3. Out-of-sample forecast on the Schiphol dataset. Actual (black) and predicted (red) values at horizon $h = 1$ for the GPD-INGARCH, GPD-INARCH, PD-INGARCH, and PD-INARCH (different plots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

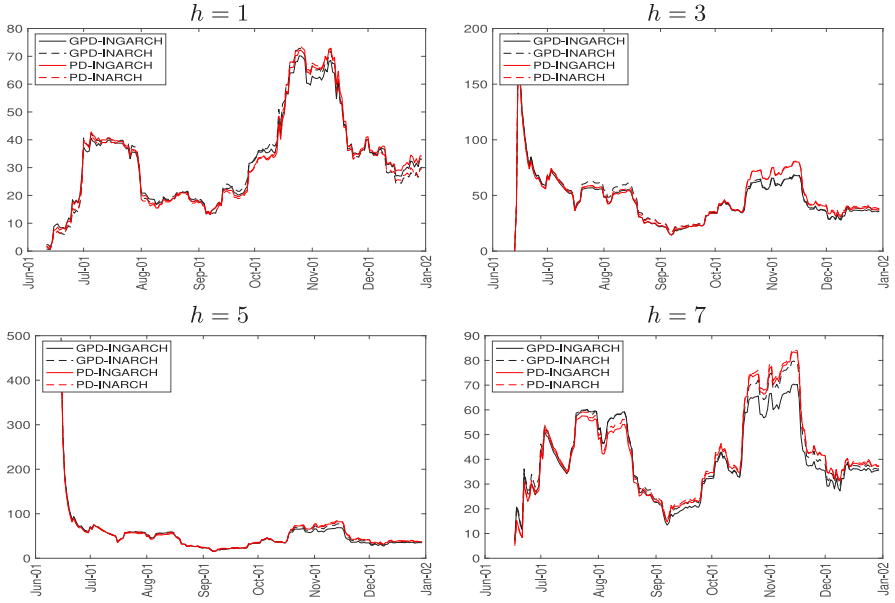

Fig. C.4. Out-of-sample forecast on the Schiphol's dataset. Mean Square Forecast Error at different horizons $h = 1, 3, 5, 7$ (different plots) on the Schiphol dataset for the GPD-INGARCH, GPD-INARCH, PD-INGARCH, and PD-INARCH (different lines).

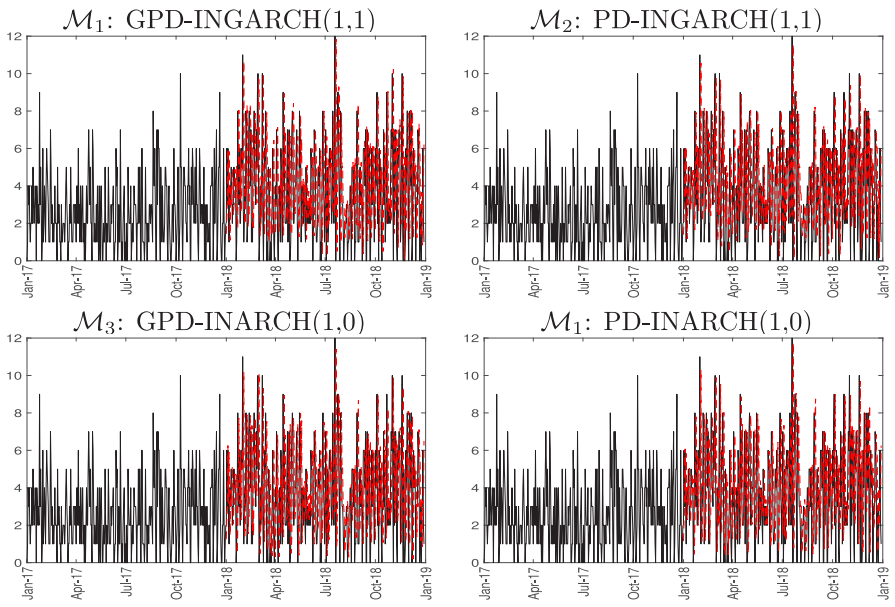

Fig. C.5. Out-of-sample forecast on the Cyber threats dataset. Actual (black) and predicted (red) values at horizon $h = 1$ for the GPD-INGARCH, GPD-INARCH, PD-INGARCH, and PD-INARCH (different plots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

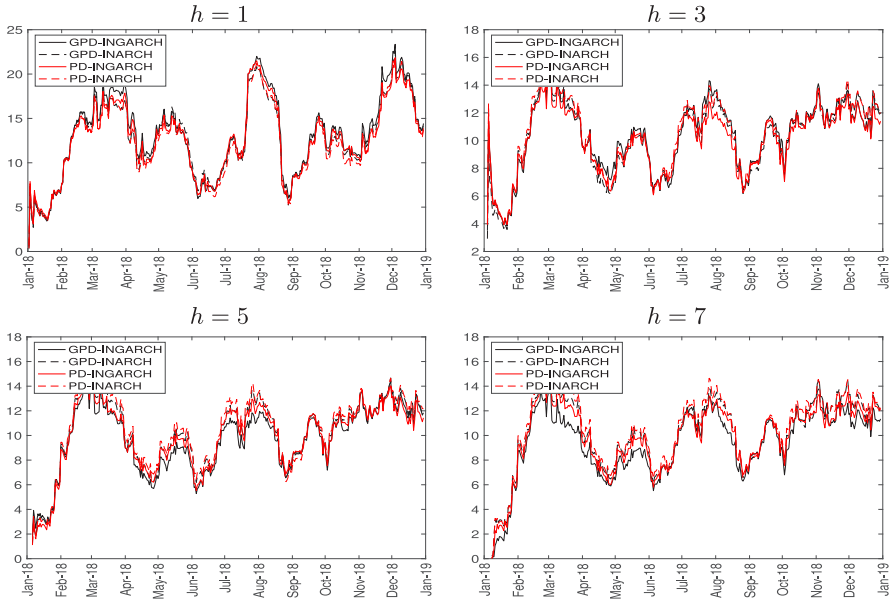

Fig. C.6. Out-of-sample forecast on the Cyber threats dataset. Mean Square Forecast Error at different horizons $h = 1, 3, 5, 7$ (different plots) on the Schiphol dataset for the GPD-INGARCH, GPD-INARCH, PD-INGARCH, and PD-INARCH (different lines).

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2023.11.009.

References

Agrafiotis, I., Nurse, J. R. C., Goldsmith, M., Creese, S., & Upton, D. (2018). A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate. *Journal of Cybersecurity*, *4*(1), tyy006.

Al-Osh, M., & Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis*, *8*(3), 261–275.

Alomani, G. A., Alzaid, A. A., Omair, M. A., et al. (2018). A Skellam GARCH model. *Brazilian Journal of Probability and Statistics*, *32*(1), 200–214.

Alzaid, A., & Al-Osh, M. (1993). Generalized Poisson ARMA processes. *Annals of the Institute of Statistical Mathematics*, *45*(2), 223–232.

Alzaid, A. A., & Omair, M. A. (2014). Poisson difference integer valued autoregressive model of order one. *Bulletin of the Malaysian Mathematical Sciences Society*, *37*(2), 465–485.

Anderson, R., & Moore, T. (2006). The economics of information security. *Science*, *314*(5799), 610–613.

Andersson, J., & Karlis, D. (2014). A parametric time series model with covariates for integers in Z. *Statistical Modelling*, *14*(2), 135–156.

Ardia, D. (2008). Bayesian estimation of a Markov-switching threshold asymmetric GARCH model with Student-t innovations. *The Econometrics Journal*, *12*(1), 105–126.

Billio, M., Casarin, R., Ravazzolo, F., & Van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, *177*(2), 213–232.

Brenner, S. W. (2004). Cybercrime metrics: Old wine, new bottles? *Virginia Journal of Law and Technology*, *9*(13), 1–53.

Brijs, T., Karlis, D., & Wets, G. (2008). Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accident Analysis and Prevention*, *40*(3).

Brockwell, P. J., Davis, R. A., & Fienberg, S. E. (1991). *Time series: Theory and methods: Theory and methods*. Springer Science & Business Media.

Cameron, E., & Pettitt, A. (2014). Recursive pathways to marginal likelihood estimation with prior-sensitivity analysis. *Statistical Science*, *29*(3), 397–419.

Cardinal, M., Roy, R., & Lambert, J. (1999). On the application of integer-valued time series models for the analysis of disease incidence. *Statistics in Medicine*, *18*(15), 2025–2039.

Celeux, G., Forbes, F., Robert, C., & Titterington, D. (2006). Foundations of DIC. *Bayesian Analysis*, *1*(4), 701–706.

Chen, C. W., & Lee, S. (2016). Generalized Poisson autoregressive models for time series of counts. *Computational Statistics & Data Analysis*, *99*, 51–67.

Chen, C. W., So, M. K., Li, J. C., & Sriboonchitta, S. (2016). Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Statistical Methodology*, *31*, 73–90.

Chib, S., Nardari, F., & Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, *108*(2), 281–316.

Consul, P. C. (1986). On the differences of two generalized Poisson variates. *Communications in Statistics – Simulation and Computation*, *15*(3), 761–767.

Cunha, E. T. d., Vasconcellos, K. L., & Bourguignon, M. (2018). A skew integer-valued time-series process with generalized Poisson difference marginal distribution. *Journal of Statistical Theory and Practice*, *12*(4), 718–743.

Davis, R. A., Dunsmuir, W. T., & Wang, Y. (1999). Modeling time series of count data. In *Statistics textbooks and monographs: vol. 158*, (pp. 63–114). MARCEL DEKKER AG.

Edwards, B., Hofmeyr, S. A., & Forrest, S. (2015). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, *2*, 3–14.

EIOPA (2019). Cyber risk for insurers – challenges and opportunities. Available at https://eiopa.europa.eu/Publications/Reports/EIOPA_Cyber_risk_for_insurers_Sept2019.pdf.

ENISA (2018). Reference incident classification taxonomy. Available at https://www.enisa.europa.eu/publications/reference-incident-classification-taxonomy.pdf.

Famoye, F. (1997). Generalized Poisson random variate generation. *American Journal of Mathematical and Management Sciences*, *17*(3–4), 219–237.

Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, *27*(6), 923–942.

Francq, C., & Zakoian, J. M. (2019). *GARCH models: Structure, statistical inference and financial applications*. Wiley.

Freeland, R. K. (1998). *Statistical analysis of discrete time series with application to the analysis of workers' compensation claims data* (Ph.D. thesis), University of British Columbia.

Freeland, R. K. (2010). True integer value time series. *AStA. Advances in Statistical Analysis*, *94*(3), 217–229.

Freeland, R., & McCabe, B. P. (2004). Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, *25*(5), 701–722.

FSB (2018). Cyber lexicon. Available at https://www.fsb.org/wp-content/uploads/P121118-1.pdf.

Gelman, A., & Meng, X. L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, *13*(2), 163–185.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 4* (pp. 169–193). Oxford: Oxford University Press.

Geweke, J. (2004). Getting it right: Joint distribution tests of posterior simulators. *Journal of the American Statistical Association*, *99*(467), 799–804.

Geyer, C. J. (1994). *Estimating normalizing constants and reweighting mixtures: Technical report 568*, School of Statistics, University of Minnesota.

Hassanien, A. E., Fouad, M. M., Manaf, A. A., Zamani, M., Ahmad, R., & Kacprzyk, J. (2016). *Multimedia forensics and security: Foundations, innovations, and applications, vol. 115*. Springer.

Husák, M., Komárková, J., Bou-Harb, E., & Čeleda, P. (2018). Survey of attack projection, prediction, and forecasting in cyber security. *IEEE Communications Surveys & Tutorials*, *21*(1), 640–660.

Jin-Guan, D., & Yuan, L. (1991). The integer-valued autoregressive (INAR (p)) model. *Journal of Time Series Analysis*, *12*(2), 129–142.

Karlis, D., & Ntzoufras, I. (2006). Bayesian analysis of the differences of count data. *Statistics in Medicine*, *25*(11), 1885–1905.

Kim, H. Y., & Park, Y. (2008). A non-stationary integer-valued autoregressive model. *Statistical Papers*, *49*(3), 485.

Koopman, S. J., Lit, R., & Lucas, A. (2017). Intraday stochastic volatility in discrete price changes: The dynamic Skellam model. *Journal of the American Statistical Association*, *112*(520), 1490–1503.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, *19*(4), 439–455.

Liesenfeld, R., Nolte, I., & Pohlmeier, W. (2006). Modelling financial transaction price movements: A dynamic integer count data model. *Empirical Economics*, *30*(4), 795–825.

Llorente, F., Martino, L., Delgado, D., & Lopez-Santiago, J. (2023). Marginal likelihood computation for model selection and hypothesis testing: An extensive review. *SIAM Review*, *65*(1), 3–58.

McAlinn, K., & West, M. (2019). Dynamic Bayesian predictive synthesis in time series forecasting. *Journal of Econometrics*, *210*(1), 155–169.

McCabe, B., & Martin, G. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting*, *21*(2), 315–330.

McCabe, B. P. M., Martin, G. M., & Harris, D. (2011). Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society. Series B*, *73*(2), 253–272.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, *21*(4), 645–650.

McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, *18*(3), 679?705.

Osh, M. A. A., & Aly, E.-E. A. (1992). First order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics – Theory and Methods*, *21*(9), 2483–2492.

Passeri, P. (2019). Hackmageddon – Information security timelines and statistics. https://www.hackmageddon.com/.

Pedeli, X., & Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical Modelling*, *11*(4), 325–349.

Robert, C., & Casella, G. (2013). *Monte carlo statistical methods*. Springer Science & Business Media.

Roberts, G. O., Gelman, A., Gilks, W. R., et al. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, *7*(1), 110–120.

Rydberg, T. H., & Shephard, N. (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, *1*(1), 2–25.

Scotto, M. G., Weiß, C. H., & Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: A review. *Statistical Modelling*, *15*(6), 590–618.

Shahtahmassebi, G., & Moyeed, R. (2014). Bayesian modelling of integer data using the generalised Poisson difference distribution. *International Journal of Statistics and Probability*, *3*(1), 35.

Shahtahmassebi, G., & Moyeed, R. (2016). An application of the generalized Poisson difference distribution to the Bayesian modelling of football scores. *Statistica Neerlandica*, *70*(3), 260–273.

Steutel, F. W., & van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, *7*(5), 893–899.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts? A survey. *Advances in Statistical Analysis*, *92*(3), 319.

Weiß, C. H. (2009). Modelling time series of counts with overdispersion. *Statistical Methods & Applications*, *18*(4), 507–519.

Werner, G., Yang, S., & McConky, K. (2017). Time series forecasting of cyber attack intensity. In *Proceedings of the 12th annual conference on cyber and information security research* (p. 18). ACM.

Xu, M., Hua, L., & Xu, S. (2017). A vine copula model for predicting the effectiveness of cyber defense early-warning. *Technometrics*, *59*(4), 508–520.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, *75*(4), 621–629.

Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, *389*(1), 58–71.

Zhu, F. K., & Li, Q. (2009). Moment and Bayesian estimation of parameters in the INGARCH(1, 1) model. *Journal of Jilin University*, *47*, 899–902.