

Generalized Poisson Difference Autoregressive Processes*

Giulia Carallo^{†§} Roberto Casarin[†] Christian P. Robert^{‡#}

[†]Ca' Foscari University of Venice, Italy

[‡]Université Paris-Dauphine, France

[#]University of Warwick, United Kindom

Abstract. This paper introduces a new stochastic process with values in the set \mathbb{Z} of integers with sign. The increments of process are Poisson differences and the dynamics has an autoregressive structure. We study the properties of the process and exploit the thinning representation to derive stationarity conditions and the stationary distribution of the process. We provide a Bayesian inference method and an efficient posterior approximation procedure based on Monte Carlo. Numerical illustrations on both simulated and real data show the effectiveness of the proposed inference.

Keywords: Bayesian inference, Counts time series, Cyber risk, Poisson Processes.

MSC2010 subject classifications. Primary 62G05, 62F15, 62M10, 62M20.

1 Introduction

In many real-world applications, time series of counts are commonly observed given the discrete nature of the variables of interest. Integer-valued variables appear very frequently in many fields, such as medicine (see Cardinal et al. (1999)), epidemiology (see Zeger (1988) and Davis et al. (1999)), finance (see Liesenfeld et al. (2006) and Rydberg and Shephard (2003)), economics (see Freeland (1998) and Freeland and McCabe (2004)), in social sciences (see Pedeli and Karlis (2011)),

*Authors' research used the SCSCF and HPC multiprocessor cluster systems at University Ca' Foscari. This work was funded in part by the Université Franco-Italienne "Visiting Professor Grant" UIF 2017-VP17_12, by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and by the Institut Universitaire de France.

[§]Corresponding author: giulia.carallo@unive.it (G. Carallo). Other contacts: r.casarin@unive.it (R. Casarin). Christian.Robert@ceremade.dauphine.fr (C. P. Robert).

sports (see Shahtahmassebi and Moyeed (2016)) and oceanography (see Cunha et al. (2018)). In this paper, we build on Poisson models, which is one of the most used model for counts data and propose a new model for integer-valued data with sign based on the generalized Poisson difference (GPD) distribution. An advantage in using this distribution relies on the possibility to account for overdispersed data with more flexibility, with respect to the standard Poisson difference distribution, a.k.a. Skellam distribution. Despite of its flexibility, GPD models have not been investigated and applied to many fields, yet. Shahtahmassebi and Moyeed (2014) proposed a GPD distribution obtained as the difference of two underlying generalized Poisson (GP) distributions with different intensity parameters. They showed that this distribution is a special case of the GPD by Consul (1986) and studied its properties. They provided a Bayesian framework for inference on GPD and a zero-inflated version of the distribution to deal with the excess of zeros in the data. Shahtahmassebi and Moyeed (2016) showed empirically that GPD can perform better than the Skellam model.

As regards to the construction method, two main classes of models can be identified in the literature: parameter driven and observation driven. In parameter-driven models the parameters are functions of an unobserved stochastic process, and the observations are independent conditionally on the latent variable. In the observation-driven models the parameter dynamics is a function of the past observations. Since this paper focuses on the observation-driven approach, we refer the reader to MacDonald and Zucchini (1997) for a review of parameter-driven models.

Thinning operators are a key ingredient for the analysis of observation-driven models. The mostly used thinning operator is the binomial thinning, introduced by Steutel and van Harn (1979) for the definition of self-decomposable distribution for positive integer-valued random variables. In mathematical biology, the binomial thinning can be interpreted as natural selection or reproduction, and in probability it is widely applied to study integer-valued processes. The binomial thinning has been generalized along different directions. Latour (1998) proposed a generalized binomial thinning where individuals can reproduce more than once. Kim and Park (2008) introduced the signed binomial thinning, in order to allow for negative values. Joe (1996) and Zheng et al. (2007) introduced the random coefficient thinning to account for external factors that may affect the coefficient of the thinning operation, such as unobservable environmental factors or states of the economy. When the coefficient follows a beta distribution one obtains the beta-binomial thinning (McKenzie (1985), McKenzie (1986) and Joe (1996)). Al-Osh and Aly (1992), proposed the iterated thinning, which can be used when the process has negative-binomial marginals. Alzaid and Al-Osh (1993) introduced the quasi-binomial thinning, that is more suitable for generalized Poisson processes. Zhang et al. (2010) introduced the signed generalized power series thinning operator, as a generalization of Kim and Park

(2008) signed binomial thinning. Thinning operation 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)). For a detailed review of the thinning operations and their properties different surveys can be consulted: MacDonald and Zucchini (1997), Kedem and Fokianos (2005), McKenzie (2003), Weiß (2008), Scotto et al. (2015). In this paper, we apply the quasi-binomial thinning difference.

In the integer-valued autoregressive process literature, thinning operations have been used either to define a process, such as in the literature on integer-valued autoregressive-moving average models (INARMA), or to study the properties of a process, such as in the literature on integer-valued GARCH (INGARCH). INARMA have been firstly introduced by McKenzie (1986) and Al-Osh and Alzaid (1987) by using the binomial thinning operator. Jin-Guan and Yuan (1991) extended to the higher order p the first-order INAR model of Al-Osh and Alzaid (1987). Kim and Park (2008) introduced an integer-valued autoregressive process with signed binomial thinning operator, $\text{INARS}(p)$, able for time series defined on \mathbb{Z} . Andersson and Karlis (2014) introduced SINARS, that is a special case of INARS model with Skellam innovations. In order to allow for negative integers, Freeland (2010) proposed a true integer-valued autoregressive model ($\text{TINAR}(1)$), that can be seen as the difference between two independent Poisson INAR(1) process. Alzaid and Al-Osh (1993) have studied an integer-valued ARMA process with Generalized Poisson marginals while Alzaid and Omair (2014) proposed a Poisson difference INAR(1) model. Cunha et al. (2018) firstly applied the GPD distribution to build a stochastic process. The authors proposed an INAR with GPD marginals and provided the properties of the process, such as mean, variance, kurtosis and conditional properties.

Rydberg and Shephard (2000) introduced heteroskedastic integer-valued processes with Poisson marginals. Later on, Heinen (2003) introduced an autoregressive conditional Poisson model and Ferland et al. (2006) proposed the INGARCH process. Both models have Poisson margins. Zhu (2012) defined a INGARCH process to model overdispersed and underdispersed count data with GP margins and Alomani et al. (2018) proposed a Skellam model with GARCH dynamics for the variance of the process. Koopman et al. (2014) proposed a Generalized Autoregressive Score (GAS) Skellam model. In this paper, we extend Ferland et al. (2006) and Zhu (2012) by assuming GPD marginals for the INGARCH model, and use the quasi-binomial thinning difference to study the properties of the new process.

Another contribution of the paper regards the inference approach. In the literature, maximum likelihood estimation has been widely investigated for integer-valued processes, whereas a very few papers discuss Bayesian inference procedures. Chen and Lee (2016) introduced Bayesian zero-inflated GP-INGARCH, with structural breaks. Zhu and Li (2009) proposed a Bayesian Poisson INGARCH(1,1)

and Chen et al. (2016) a Bayesian Autoregressive Conditional Negative Binomial model. In this paper, we develop a Bayesian inference procedure for the proposed GPD-INGARCH process and a Markov chain Monte Carlo (MCMC) procedure for posterior approximation. One of the advantages of the Bayesian approach is that extra-sample information on the parameters value can be included in the estimation process through the prior distributions. Moreover, it can be easily combined with a data augmentation strategy to make the likelihood function more tractable.

We apply our model to a cyber-threat dataset and contribute to cyber-risk literature providing evidence of temporal patterns in the mean and variance of the threats, which can be used to predict threat arrivals. Cyber threats are increasingly considered as a top global risk for the financial and insurance sectors and for the economy as a whole (e.g. EIOPA, 2019). As pointed out in Hassanien et al. (2016), the frequency of cyber events substantially increased in the past few years and cyber-attacks occur on a daily basis. Understanding cyber-threats dynamics and their impact is critical to ensure effective controls and risk mitigation tools. Despite these evidences and the relevance of the topic, the research on the analysis of cyber threats is scarce and scattered in different research areas such as cyber security (Agrafiotis et al., 2018), criminology Brenner (2004), economics Anderson and Moore (2006) and sociology. In statistics there are a few works on modelling and forecasting cyber-attacks. Xu et al. (2017) introduced a copula model to predict effectiveness of cyber-security. Werner et al. (2017) used an autoregressive integrated moving average model to forecast the number of daily cyber-attacks. Edwards et al. (2015) apply Bayesian Poisson and negative binomial models to analyse data breaches and find evidence of over-dispersion and absence of time trends in the number of breaches. See Husák et al. (2018) for a review on modelling cyber threats.

The paper is organized as follows. In Section 2 we introduce the parametrization used for the GPD and define the GPD-INGARCH process. Section 3 aims at studying the properties of the process. Section 4 presents a Bayesian inference procedure. Section 5 and 6 provide some illustration on simulated and real data, respectively. Section 7 concludes.

2 Generalized Poisson Difference INGARCH

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

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

with parameters $\theta > 0$ and $0 \leq \lambda < 1$ (see Consul, 1986). We denote this distribution with $GP(\theta, \lambda)$. Let $X \sim GP(\theta_1, \lambda)$ and $Y \sim GP(\theta_2, \lambda)$ be two independent

GP random variables, Consul (1986) showed that the probability distribution of $Z = (X - Y)$ follows a Generalized Poisson Difference distribution (GPD) with pmf:

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

where z takes integer values in the interval $(-\infty, +\infty)$ and $0 < \lambda < 1$ and $\theta_1, \theta_2 > 0$ are the parameters of the distribution. See Appendix A.3 for a more general definition of the GPD with possibly negative λ .

In the following Lemma we state the convolution property of the GPD distribution since which will be used in this paper. Appendix B.1 provides an original proof of this result.

Lemma 1 (Convolution Property). *The sum of two independent random GPD variates, $X + Y$, with parameters $(\theta_1, \theta_2, \lambda)$ and $(\theta_3, \theta_4, \lambda)$ is a GPD variate with parameters $(\theta_1 + \theta_3, \theta_2 + \theta_4, \lambda)$. The difference of two independent random GPD variates, $X - Y$, with parameters $(\theta_1, \theta_2, \lambda)$ and $(\theta_3, \theta_4, \lambda)$ is a GPD variate with parameters $(\theta_1 + \theta_4, \theta_2 + \theta_3, \lambda)$.*

We use an equivalent pmf and a re-parametrization of the GPD, which are better suited for the definition of a INGARCH model. A random variable Z follows a GPD if and only if its probability distribution is

$$P_z(\mu, \sigma^2, \lambda) = e^{-\sigma^2 - z\lambda} \sum_{s=\max(0, -z)}^{+\infty} \frac{1}{4} \frac{\sigma^4 + \mu^2}{s!(s+z)!} \left[\frac{\sigma^2 + \mu}{2} + (s+z)\lambda \right]^{s+z-1} \left[\frac{\sigma^2 - \mu}{2} + s\lambda \right]^{s-1} e^{-2\lambda s} \quad (3)$$

We denote this distribution with $GPD(\mu, \sigma^2, \lambda)$.

Remark 1. *The probability distribution in Eq. 3 is equivalent to the one in Eq. 2 up to the reparametrization $\mu = \theta_1 - \theta_2$ and $\sigma^2 = \theta_1 + \theta_2$. See Appendix B for a proof.*

The mean, variance, skewness and kurtosis of a GPD random variable can be obtained in close form by exploiting the representation of the GPD as difference between independent GP random variables.

Remark 2. *Let $Z \sim GPD(\mu, \sigma^2, \lambda)$, then mean and variance are:*

$$E(Z) = \frac{\mu}{1 - \lambda}, \quad V(Z) = \frac{\sigma^2}{(1 - \lambda)^3} \quad (4)$$

and the Pearson skewness and kurtosis are:

$$S(Z) = \frac{\mu}{\sigma^3} \frac{(1 + 2\lambda)}{\sqrt{1 - \lambda}}, \quad K(Z) = 3 + \frac{1 + 8\lambda + 6\lambda^2}{\sigma^2(1 - \lambda)} \quad (5)$$

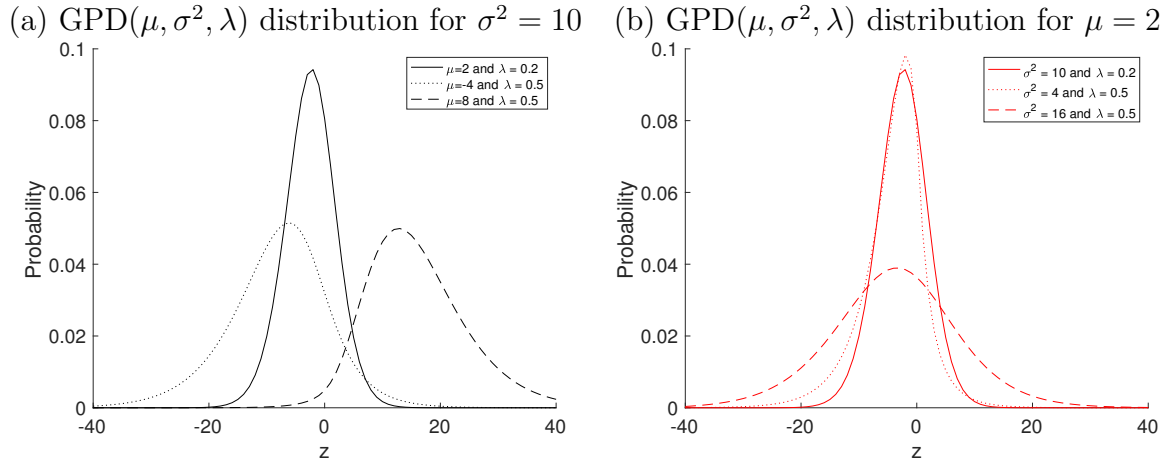


Figure 1: Generalized Poisson difference distribution $GPD(\mu, \sigma^2, \lambda)$ for some values of λ , μ and σ^2 . The distribution with $\lambda = 0.2$, $\mu = 2$ and $\sigma^2 = 10$ (solid line) is taken as baseline in both panels.

See Appendix B for a proof.

Figure 1 shows the sensitivity of the probability distribution with respect to the location parameter μ (panel a), the scale parameter σ^2 (panel b) and the skewness parameter λ (different lines in each plot). For given values of λ and μ , when σ^2 decreases the dispersion of the GPD increases (dotted and dashed lines in the 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 in the left plot). See Appendix A.3 for further numerical illustrations.

Differently from the usual $GARCH(p, q)$ process (e.g., see Francq and Zakoian (2019)), the $INGARCH(p, q)$ 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

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

with

$$\frac{\tilde{\mu}_t}{1 - \lambda} = \mu_t = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i} + \sum_{j=1}^q \beta_j \mu_{t-j} \quad (6)$$

where $\tilde{\mu}_{t-j} = \mu_{t-j}(1 - \lambda)$, $\alpha_0 \in \mathbb{R}$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $i = 1, \dots, p$, $p \geq 1$, $j = 1, \dots, q$, $q \geq 0$. For $q = 0$ the model reduces to a $GPD-INGARCH(p)$ and for $\lambda = 0$ one obtains a Skellam $INGARCH(p, q)$ which extends to Poisson differences the Poisson

INGARCH(p, q) of Ferland et al. (2006). From the properties of the GPD, the conditional mean $\mu_t = E(Z_t | \mathcal{F}_{t-1})$ and variance $\sigma_t^2 = V(Z_t | \mathcal{F}_{t-1})$ of the process are:

$$\mu_t = \frac{\tilde{\mu}_t}{1 - \lambda}, \quad \sigma_t^2 = \frac{\tilde{\sigma}_t^2}{(1 - \lambda)^3} \quad (7)$$

respectively. In the application, we assume $\sigma_t^2 = |\mu_t|\phi$. Following the parametrization defined in Remark 1, we need to impose the constrain $\phi > (1 - \lambda)^{-2}$, in order to have a well-defined GPD distribution. In Fig. 2, we provide some simulated examples of the GPD-INGARCH(1, 1) process for different values of the parameters α_0 , α_1 and β_1 .

Simulations from a GPD-INGARCH are obtained as differences of GP sequences

$$Z_t = X_t - Y_t, \quad X_t \sim GP(\theta_{1t}, \lambda), \quad Y_t \sim GP(\theta_{2t}, \lambda)$$

where

$$\theta_{1t} = \frac{\sigma_t^2 + \mu_t}{2}, \quad \theta_{2t} = \frac{\sigma_t^2 - \mu_t}{2}. \quad (8)$$

Each random sequence is generated by the branching method in Famoye (1997), which performs faster than the inversion method for large values of θ_{1t} and θ_{2t} . We considered 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 column in Fig. 2). The first and second line show paths for positive and negative value of the intercept α_0 , respectively. The last line illustrates the effect of λ on the trajectories with respect to the baselines in Panels (a) and (b). Comparing (I.a) and (I.b) in Fig. 3 one can see that increasing β_1 increases serial correlation and the kurtosis level (compare (II.a) and (II.b)).

We provide a necessary condition on the parameters α_i and β_j that will ensure that a second-order stationary process has an INGARCH representation. First define the two following polynomials: $D(B) = 1 - \beta_1 B - \dots - \beta_q B^q$ and $G(B) = \alpha_1 B + \dots + \alpha_p B^p$, where B is the backshift operator. Assume the roots of $D(z)$ lie outside the unit circle. For non-negative β_j this is equivalent to assume $D(1) = \sum_{j=1}^q \beta_j < 1$. Then, the operator $D(B)$ has inverse $D^{-1}(B)$ and it is possible to write

$$\mu_t = D^{-1}(B)(\alpha_0 + G(B)Z_t) = \alpha_0 D^{-1}(1) + H(B)Z_t \quad (9)$$

where $H(B) = G(B)D^{-1}(B) = \sum_{j=1}^{\infty} \psi_j B^j$ and ψ_j are given by the power expansion of the rational function $G(z)/D(z)$ in the neighbourhood of zero. If we denote $K(B) = D(B) - G(B)$ we can write the necessary condition as in the following proposition.

Proposition 1. *A necessary condition for a second-order stationary process $\{Z_t\}_{t \in \mathbb{Z}}$ to satisfy Eq. 6 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 B □

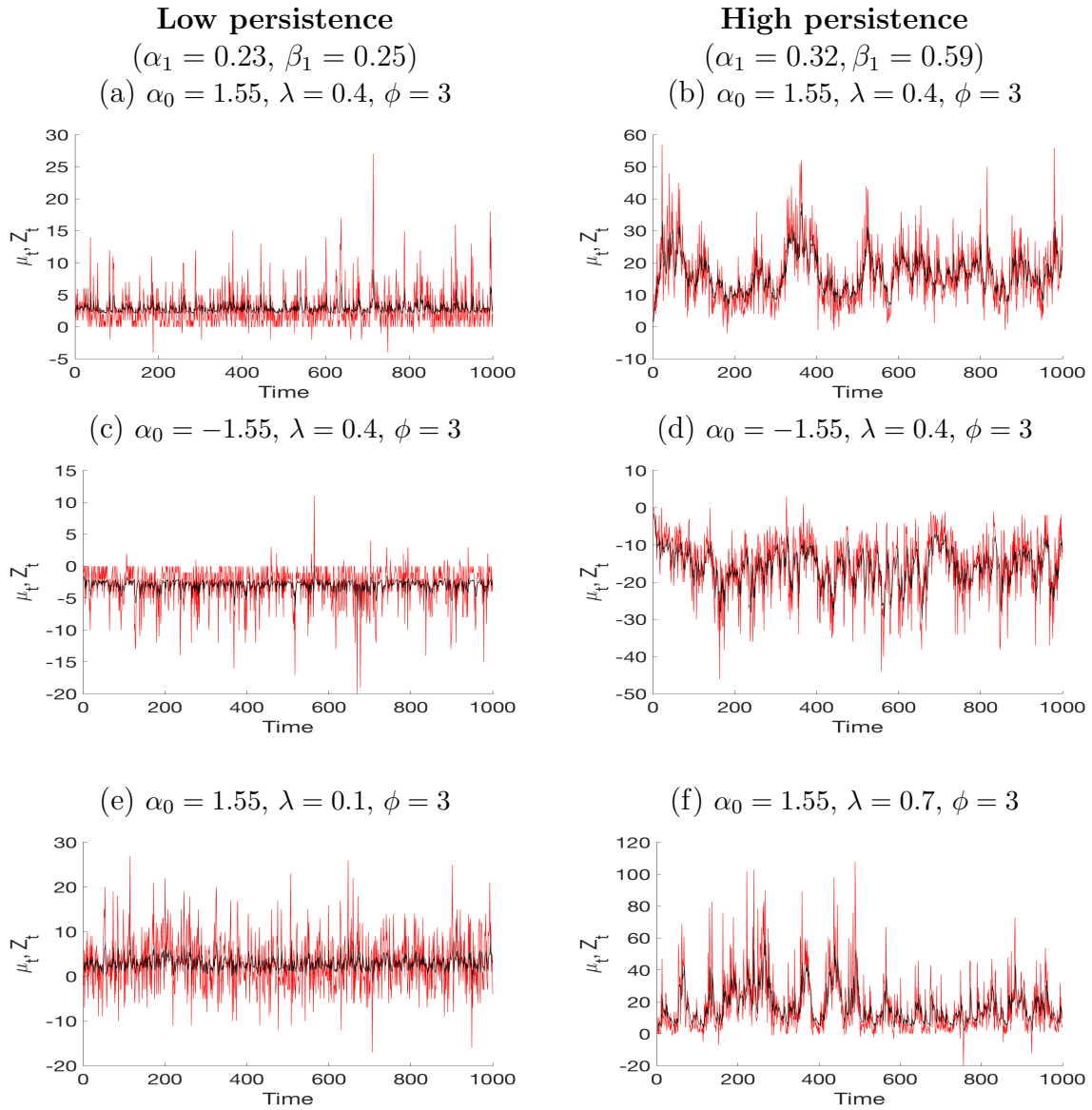


Figure 2: Simulated INGARCH(1,1) paths for different values of the parameters α_0 , α_1 and β_1 . In Panels from (a) to (d) the effect of α_0 ($\alpha_0 > 0$ in the first line and $\alpha_0 < 0$ in the second line) with $\lambda = 0.4$ and $\phi = 3$. In Panels (e) and (f) the effect of lambda ($\lambda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

3 Properties of the GPD-INGARCH

We study the properties of the process by exploiting a suitable thinning representation following the strategy in Ferland et al. (2006) and Zhu (2012) for

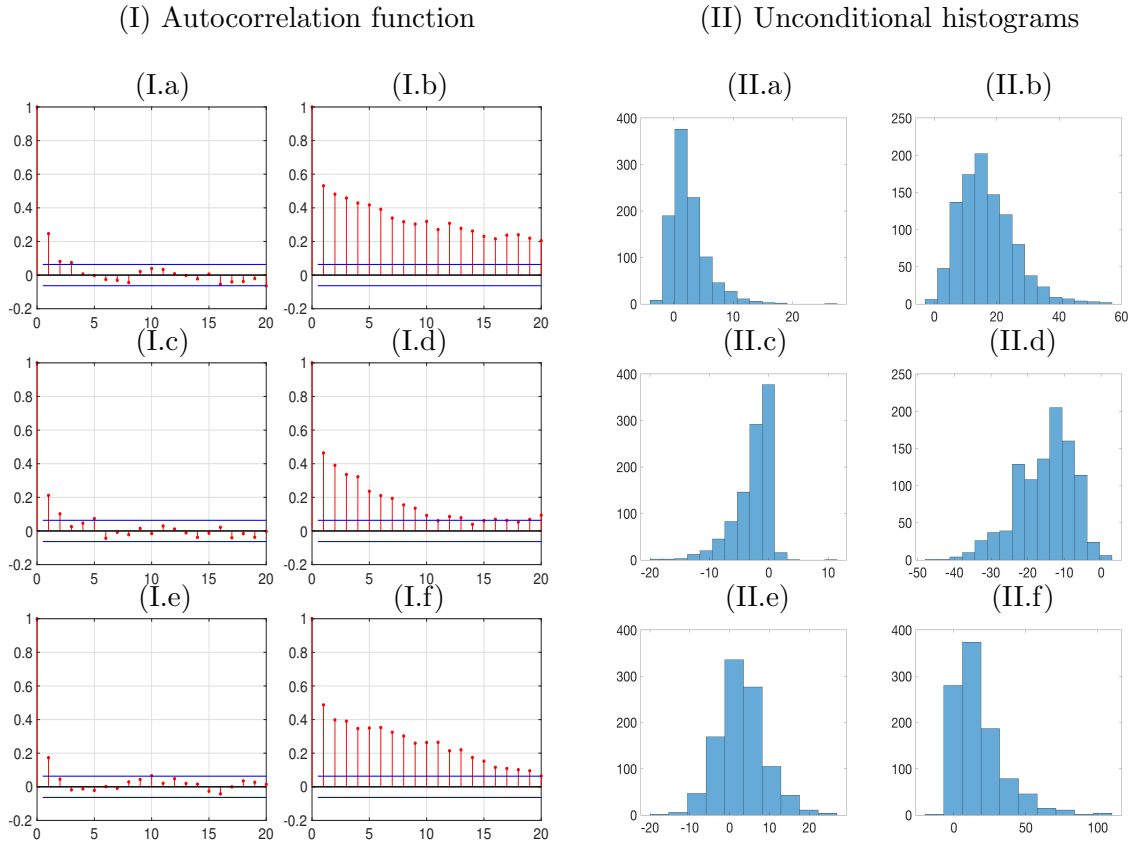


Figure 3: Autocorrelation functions (Panel I) and unconditional distributions (Panel II) of $\{Z_t\}_{t \in \mathbb{Z}}$ for the different cases presented in Fig. 2 (different columns in each panel).

Poisson 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.

3.1 Thinning representation

We show that the INGARCH process can be obtained as a limit of successive approximations. Let us define:

$$X_t^{(n)} = \begin{cases} 0, & n < 0 \\ (1 - \lambda)U_{1t}, & n = 0 \\ (1 - \lambda)U_{1t} + (1 - \lambda) \sum_{i=1}^n \sum_{j=1}^{\frac{X_{t-i}^{(n-i)}}{(1-\lambda)}} V_{1t-i,i,j}, & n > 0 \end{cases} \quad (10)$$

and

$$Y_t^{(n)} = \begin{cases} 0, & n < 0 \\ (1 - \lambda)U_{2t}, & n = 0 \\ (1 - \lambda)U_{2t} + (1 - \lambda) \sum_{i=1}^n \sum_{j=1}^{\frac{Y_{t-i}^{(n-i)}}{(1-\lambda)}} V_{2t-i,i,j}, & n > 0 \end{cases} \quad (11)$$

where $\{U_{1t}\}_{t \in \mathbb{Z}}$ and $\{U_{2t}\}_{t \in \mathbb{Z}}$ are sequences of independent GP random variables and for each $t \in \mathbb{Z}$ and $i \in \mathbb{N}$, $\{V_{1t,i,j}\}_{j \in \mathbb{N}}$ and $\{V_{2t,i,j}\}_{j \in \mathbb{N}}$ represent two sequences of independent integer random variables. Moreover, assume that all the variables U_s , $V_{t,i,j}$, with $s \in \mathbb{Z}$, $t \in \mathbb{Z}$, $i \in \mathbb{N}$ and $j \in \mathbb{N}$, are mutually independent.

It is possible to show that $X_t^{(n)}$ and $Y_t^{(n)}$ have a thinning representation. We define a suitable thinning operation, first used by Alzaid and Al-Osh (1993) and follow the notation in Weiß (2008), let $\rho_{\theta,\lambda} \circ$ be the quasi-binomial thinning operator, such that it follows a QB($\rho, \theta/\lambda, x$).

Proposition 2. *If X follows a GP(λ, θ) distribution and the 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). □

Both $X_t^{(n)}$ and $Y_t^{(n)}$ in Eq. 10 and 11 admit the representation

$$X_t^{(n)} = (1 - \lambda)U_{1t} + (1 - \lambda) \sum_{i=1}^n \varphi_{1i}^{(t-i)} \circ \left(\frac{X_{t-i}^{(n-i)}}{1 - \lambda} \right), \quad n > 0 \quad (12)$$

and

$$Y_t^{(n)} = (1 - \lambda)U_{2t} + (1 - \lambda) \sum_{i=1}^n \varphi_{2i}^{(t-i)} \circ \left(\frac{Y_{t-i}^{(n-i)}}{1 - \lambda} \right), \quad n > 0 \quad (13)$$

where $\varphi \circ X$ is the quasi-binomial thinning operation. See Appendix A for a definition.

In the following we introduce the thinning difference operator and show that $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$ has a thinning representation.

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

$$\rho \diamond Z | Z \stackrel{d}{=} (\rho_{\theta_1, \lambda} \circ X) - (\rho_{\theta_2, \lambda} \circ Y) | (X - Y) \quad (14)$$

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 " $A \stackrel{d}{=} B$ " means that the random variables A and B have the same distribution.

See Cunha et al. (2018) for an application of the thinning operation to GPD-INAR processes and Appendix A.4 for further details. Using the new operator as defined in Eq. 14, we can represent $Z_t^{(n)}$ as follows.

Proposition 3. *The process $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$ has the representation:*

$$Z_t^{(n)} = (1 - \lambda)U_t + (1 - \lambda)^2 \sum_{i=1}^n \varphi_i^{(t-i)} \diamond \left(\frac{Z_{t-i}^{(n-i)}}{1 - \lambda} \right), \quad n > 0 \quad (15)$$

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 $\psi_0 = \alpha_0/D(1)$.

Proof. See Appendix B □

The proposition above shows that $Z_t^{(n)}$ is obtained through a cascade of thinning operations along the sequence $\{U_t\}_{t \in \mathbb{Z}}$. For example:

$$\begin{aligned} Z_t^{(0)} &= (1 - \lambda)U_t \\ Z_t^{(1)} &= (1 - \lambda)U_t + (1 - \lambda)^2 \left[\varphi_1^{(t-1)} \diamond (Z_{t-1}^{(1-1)}/(1 - \lambda)) \right] \\ &= (1 - \lambda)U_t + (1 - \lambda)^2 (\varphi_1^{(t-1)} \diamond U_{t-1}) \\ Z_t^{(2)} &= (1 - \lambda)U_t + (1 - \lambda)^2 \left[\varphi_1^{(t-1)} \diamond (Z_{t-1}^{(2-1)}/(1 - \lambda)) + \varphi_2^{(t-2)} \diamond (Z_{t-2}^{(2-2)}/(1 - \lambda)) \right] \\ &= (1 - \lambda)U_t + (1 - \lambda)^2 \left[\varphi_1^{(t-1)} \diamond U_{t-1} + \varphi_1^{(t-1)} \diamond (\varphi_1^{(t-2)} \diamond U_{t-2}) + \varphi_2^{(t-2)} \diamond U_{t-2} \right] \\ Z_t^{(3)} &= (1 - \lambda)U_t + (1 - \lambda)^2 \left[\varphi_1^{(t-1)} \diamond (Z_{t-1}^{(3-1)}/(1 - \lambda)) + \varphi_2^{(t-2)} \diamond (Z_{t-2}^{(3-2)}/(1 - \lambda)) + \right. \\ &\quad \left. + \varphi_3^{(t-3)} \diamond (Z_{t-3}^{(3-3)}/(1 - \lambda)) \right] \\ &= (1 - \lambda)U_t + (1 - \lambda)^2 \left[\varphi_1^{(t-1)} \diamond U_{t-1} + \varphi_1^{(t-1)} \diamond (\varphi_1^{(t-2)} \diamond U_{t-2}) + \right. \\ &\quad \left. + \varphi_2^{(t-2)} \diamond U_{t-2} + \varphi_1^{(t-1)} \diamond (\varphi_1^{(t-2)} \diamond (\varphi_1^{(t-3)} \diamond U_{t-3})) + \right. \\ &\quad \left. + \varphi_1^{(t-1)} \diamond (\varphi_2^{(t-3)} \diamond U_{t-3}) + \varphi_2^{(t-2)} \diamond (\varphi_1^{(t-3)} \diamond U_{t-3}) + \varphi_3^{(t-3)} \diamond U_{t-3} \right]. \end{aligned}$$

Since $Z_t^{(n)}$ is a finite weighted sum of independent GPD random variables, the expected value and the variance of $Z_t^{(n)}$ are well defined. Moreover, it can be seen that $E[Z_t^{(n)}]$ does not depend on t but only on n , hence it can be denoted as μ_n . Using Proposition 3 and $\mu_k = 0$ if $k < 0$, it is possible to write μ_n as follows

$$\begin{aligned} \mu_n &= (1 - \lambda)E[U_t] + (1 - \lambda)^2 \sum_{i=1}^n E \left[\varphi_i^{(t-i)} \diamond \left(\frac{Z_{t-i}^{(n-i)}}{1 - \lambda} \right) \right] \\ &= \psi_0 + \sum_{j=1}^{\infty} \psi_j \mu_{n-j} = D^{-1}(B)\alpha_0 + H(B)\mu_n \end{aligned} \quad (16)$$

from which it follows $D(B)\mu_n = G(B)\mu_n + \alpha_0 \Leftrightarrow K(B)\mu_n = \alpha_0$, where $K(B) = D(B) - G(B)$. From the last equation it can be seen that the sequence $\{\mu_n\}$ 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 the assumption $K(1) > 0$, the following holds true.

Proposition 4. *If $K(1) > 0$ then the sequence $\{Z_t^{(n)}\}_{n \in \mathbb{N}}$ has an almost sure limit.*

Proof. See Appendix B. □

Proposition 5. *If $K(1) > 0$ then the sequence $\{Z_t^{(n)}\}$ has a mean-square limit.*

Proof. See Appendix B. □

3.2 Stationarity

Given Proposition 4, if we can show that $\{Z_t^{(n)}\}$ is a strictly stationary process, for any given n , then also its almost sure limit $\{Z_t\}_{t \in \mathbb{Z}}$ will be a strictly stationary process. In order to show stationarity for $\{Z_t^{(n)}\}$, we follow a procedure similar to the one in Ferland et al. (2006). Let us define the probability generating function (pgf) $g_{\mathbf{W}}(\mathbf{t})$ of the random vector $\mathbf{W} = (W_1, \dots, W_k)$

$$g_{\mathbf{W}}(\mathbf{t}) = E \left[\prod_{i=1}^k t_i^{W_i} \right] = \sum_{\mathbf{w} \in \mathbb{N}^k} p(\mathbf{w}) \prod_{i=1}^k t_i^{w_i} \quad (17)$$

where $p(\mathbf{W}) = Pr(\mathbf{W} = (W_1, \dots, W_k)')$ and $\mathbf{t} = (t_1, \dots, t_k)' \in \mathbb{C}^k$. The probability generating function has the following properties.

Proposition 6. *Let $\mathbf{Z}_{1\dots k}^{(n)} = (Z_1^{(n)}, \dots, Z_k^{(n)})$ be a subsequence of $\{Z_t^{(n)}\}_{t \in \mathbb{Z}}$ where, without loss of generality, we choose the first k periods. Let $\mathbf{X}_{1\dots k}^{(n)} = (X_1^{(n)}, \dots, X_k^{(n)})$ and $\mathbf{Y}_{1\dots k}^{(n)} = (Y_1^{(n)}, \dots, Y_k^{(n)})$ be such that $\mathbf{Z}_{1\dots k}^{(n)} = (\mathbf{X}_{1\dots k}^{(n)} - \mathbf{Y}_{1\dots k}^{(n)})'$ then*

$$g_{\mathbf{Z}_{1\dots k}^{(n)}}(\mathbf{t}) = g_{\mathbf{X}_{1\dots k}^{(n)}}(\mathbf{t})g_{\mathbf{Y}_{1\dots k}^{(n)}}(\mathbf{t}^{-1}) \quad (18)$$

Proof. See Appendix B □

Using the probability generating function, in the following we know the stationarity of the process.

Proposition 7. *$\{Z_t^{(n)}\}_{t \in \mathbb{Z}}$ is a strictly stationary process, for any fixed value of n .*

Proof. Let k and h be two positive integers. As pointed out by Ferland et al. (2006), Brockwell et al. (1991) show that to prove strictly stationarity we only need to show that

$$\mathbf{Z}_{1+h\dots k+h}^{(n)} = (Z_{1+h}^{(n)}, \dots, Z_{k+h}^{(n)})' \text{ and } \mathbf{Z}_{1\dots k}^{(n)} = (Z_1^{(n)}, \dots, Z_k^{(n)})' \quad (19)$$

have the same joint distribution, where we can rewrite both vectors in Eq. 19 as

$$\begin{aligned} \mathbf{Z}_{1+h\dots k+h}^{(n)} &= (\mathbf{X}_{1+h\dots k+h}^{(n)} - \mathbf{Y}_{1+h\dots k+h}^{(n)})' \\ &= ((X_{1+h}^{(n)} - Y_{1+h}^{(n)}), \dots, (X_{k+h}^{(n)} - Y_{k+h}^{(n)}))' \end{aligned} \quad (20)$$

and

$$\begin{aligned} \mathbf{Z}_{1\dots k}^{(n)} &= (\mathbf{X}_{1\dots k}^{(n)} - \mathbf{Y}_{1\dots k}^{(n)})' \\ &= ((X_1^{(n)} - Y_1^{(n)}), \dots, (X_k^{(n)} - Y_k^{(n)}))' \end{aligned} \quad (21)$$

To show that the two vectors have the same probability generating function, we first write the pgfs of X , Y and Z as shown above.

$$\begin{aligned} g_{\mathbf{X}_{1\dots k}^{(n)}}(\mathbf{t}) &= E \left[\prod_{j=1}^k t_j^{X_j^{(n)}} \right] = E \left[E_{\mathbf{X}_{1\dots k}^{(n)} | \mathbf{U}_{1,1-n\dots k}} \left[\prod_{j=1}^k t_j^{X_j^{(n)}} \right] \right] \\ &= \sum_{\mathbf{v}_1 \in \mathbb{N}^{(k+n)}} E_{\mathbf{X}_{1\dots k}^{(n)} | \mathbf{U}_{1,1-n\dots k} = \mathbf{v}_1} \left[\prod_{j=1}^k t_j^{X_j^{(n)}} \right] Pr(\mathbf{U}_{1,1-n\dots k} = \mathbf{v}_1) \end{aligned} \quad (22)$$

$$g_{\mathbf{Y}_{1\dots k}^{(n)}}(\mathbf{t}) = \sum_{\mathbf{v}_2 \in \mathbb{N}^{(k+n)}} E_{\mathbf{Y}_{1\dots k}^{(n)} | \mathbf{U}_{2,1-n\dots k} = \mathbf{v}_2} \left[\prod_{j=1}^k t_j^{Y_j^{(n)}} \right] Pr(\mathbf{U}_{2,1-n\dots k} = \mathbf{v}_2) \quad (23)$$

$$G_{\mathbf{Z}_{1\dots k}^{(n)}}(\mathbf{t}) = \sum_{\mathbf{v} \in \mathbb{N}^{(k+n)}} E_{\mathbf{Z}_{1\dots k}^{(n)} | \mathbf{U}_{1-n\dots k} = \mathbf{v}} \left[\prod_{j=1}^k t_j^{(X_j^{(n)} - Y_j^{(n)})} \right] Pr(\mathbf{U}_{1-n\dots k} = \mathbf{v}) \quad (24)$$

By the thinning representation, for a any given value $\mathbf{u}_{1,t-n\dots t+k} = (u_{1,t-n}, \dots, u_{1,t+k})'$ of the vector $\mathbf{U}_{1,t-n\dots t+k} = (U_{1,t-n}, \dots, U_{1,t+k})'$ and $\mathbf{u}_{2,t-n\dots t+k} = (u_{2,t-n}, \dots, u_{2,t+k})'$ of the vector $\mathbf{U}_{2,t-n\dots t+k} = (U_{2,t-n}, \dots, U_{2,t+k})'$, the components of the vectors $(X_1^{(n)}, \dots, X_k^{(n)})'$ and $(Y_1^{(n)}, \dots, Y_k^{(n)})'$ 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

$$\mathbf{Z}_{1+h\dots k+h}^{(n)} = ((X_{1+h}^{(n)} - Y_{1+h}^{(n)}), \dots, (X_{k+h}^{(n)} - Y_{k+h}^{(n)}))'$$

and

$$\mathbf{Z}_{1\dots k}^{(n)} = ((X_1^{(n)} - Y_1^{(n)}), \dots, (X_k^{(n)} - Y_k^{(n)}))'$$

given $\mathbf{U}_{t-n\dots t+k}$ and $\mathbf{U}_{t-n+h\dots t+k+h}$, are the same. Accordingly,

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

and, since

$$Pr(\mathbf{U}_{1-n+h\dots k+h} = \mathbf{v}) = Pr(\mathbf{U}_{1-n\dots k} = \mathbf{v}),$$

it is possible to write

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

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

Proposition 8. *The process $\{Z_t\}_{t \in \mathbb{Z}}$ is a strictly stationary process.*

Proposition 9. *The first two moments of $\{Z_t\}_{t \in \mathbb{Z}}$ are finite.*

Proof. See Appendix B. \square

3.3 Conditional law of $\{Z_t^{(n)}\}_{t \in \mathbb{Z}}$ given \mathcal{F}_{t-1}

To verify that the distributional properties of the sequence are satisfied, we will follow the same arguments in Ferland et al. (2006) adjusted for our sequence. Given $\mathcal{F}_{t-1} = \sigma(\{Z_u\}_{u \leq t-1})$, 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

$$\mu_t = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i} + \sum_{j=1}^q b_j \mu_{t-j}. \quad (25)$$

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

$$r_{1t}^{(n)} = (1 - \lambda)U_{1t} + (1 - \lambda) \sum_{i=1}^n \sum_{j=1}^{X_{t-i}} V_{1t-i,i,k} \quad (26)$$

$$r_{2t}^{(n)} = (1 - \lambda)U_{2t} + (1 - \lambda) \sum_{i=1}^n \sum_{j=1}^{Y_{t-i}} V_{2t-i,i,k}. \quad (27)$$

and

$$r_t^{(n)} = r_{1t}^{(n)} - r_{2t}^{(n)}. \quad (28)$$

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

$$X_t - r_{1t}^{(n)} = (X_t - X_t^{(n)}) + (X_t^{(n)} - r_{1t}^{(n)}) \quad (29)$$

and

$$Y_t - r_{2t}^{(n)} = (Y_t - Y_t^{(n)}) + (Y_t^{(n)} - r_{2t}^{(n)}). \quad (30)$$

Since $X_t^{(n)} \xrightarrow{a.s.} X_t$ and $Y_t^{(n)} \xrightarrow{a.s.} Y_t$, we know that the first term in both Eq. 29 and 30 goes to zero. Therefore, we can write

$$\begin{aligned} Z_t - r_t^{(n)} &= (X_t - Y_t) - (r_{1t}^{(n)} - r_{2t}^{(n)}) \\ &= \left[(X_t - X_t^{(n)}) - (Y_t - Y_t^{(n)}) \right] + \left[(X_t^{(n)} - r_{1t}^{(n)}) - (Y_t^{(n)} - r_{2t}^{(n)}) \right] \\ &= (Z_t - Z_t^{(n)}) + \left[(X_t^{(n)} - Y_t^{(n)}) - (r_{1t}^{(n)} - r_{2t}^{(n)}) \right] \\ &= (Z_t - Z_t^{(n)}) + \left[Z_t^{(n)} - (r_{1t}^{(n)} - r_{2t}^{(n)}) \right], \end{aligned} \quad (31)$$

and, as before, $(Z_t - Z_t^{(n)})$ 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. 31 goes to zero, for this purpose we need to find a sequence

$$W_t^{(n)} = (r_{1t}^{(n)} - r_{2t}^{(n)}) - Z_t^{(n)}$$

that converges almost surely to zero. For this reason it is more suitable to rewrite the previous sequence as follows

$$\begin{aligned} W_t^{(n)} &= (r_{1t}^{(n)} - r_{2t}^{(n)}) - (X_t - Y_t) \\ &= (r_{1t}^{(n)} - X_t) - (r_{2t}^{(n)} - Y_t) \end{aligned} \quad (32)$$

Ferland et al. (2006) show that

$$\begin{aligned} \lim_{n \rightarrow \infty} E \left[(r_{1t}^{(n)} - X_t) \right] &= 0 \\ \lim_{n \rightarrow \infty} E \left[(r_{2t}^{(n)} - Y_t) \right] &= 0 \end{aligned} \quad (33)$$

therefore, we can conclude that also

$$\lim_{n \rightarrow \infty} E \left[(r_t^{(n)} - Z_t) \right] = 0. \quad (34)$$

Equation 34 implies that $W_t^{(n)}$ converges to zero in L^1 , therefore there exist a subsequence $W_t^{(n_k)}$ converging almost surely to the same limit. From this it follows directly that the distributional properties of X_t are satisfied.

Since $r_{1t}^{(n_k)} \xrightarrow{a.s.} X_t$ and $r_{2t}^{(n_k)} \xrightarrow{a.s.} Y_t$, it is also true $r_t^{(n_k)} \xrightarrow{a.s.} Z_t$. Hence,

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

However,

$$r_t^{(n)} | \mathcal{F}_{t-1} = (r_{1t}^{(n)} - r_{2t}^{(n)}) | \mathcal{F}_{t-1}$$

and from Zhu (2012) we know that both $r_{1t}^{(n)}$ and $r_{2t}^{(n)}$ have a Generalized Poisson distribution. Since the difference of two GP distributed random variables is GPD distributed, we can write

$$r_t^{(n)} | \mathcal{F}_{t-1} \sim GPD \left(\alpha_0 D^{-1}(1) + \sum_{j=1}^n \psi_j Z_{t-j} \right) \quad (35)$$

and conclude that

$$Z_t | \mathcal{F}_{t-1} \sim GPD(\tilde{\mu}_t, \tilde{\sigma}_t^2, \lambda). \quad (36)$$

3.4 Moments of the GPD-INGARCH

The conditional mean and variance of the process Z_t are

$$\begin{aligned} E(Z_t | \mathcal{F}_{t-1}) &= \frac{\tilde{\mu}_t}{1 - \lambda} = \mu_t \\ V(Z_t | \mathcal{F}_{t-1}) &= \frac{\tilde{\sigma}_t^2}{1 - \lambda} = \phi^3 \tilde{\sigma}_t^2 \end{aligned} \quad (37)$$

where $\phi = \frac{1}{1-\lambda}$.

The unconditional mean and variance of the process are

$$\begin{aligned} E(Z_t) &= \mu_t = \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j} \\ V(Z_t) &= E[V(Z_t | \mathcal{F}_{t-1})] + V[E(Z_t | \mathcal{F}_{t-1})] \\ &= E(\phi^3 \tilde{\sigma}_t^2) + V(\mu_t) \\ &= \phi^3 E(\tilde{\sigma}_t^2) + V(\mu_t) \end{aligned} \quad (38)$$

From Th. 1 in Weiß (2009) we know a set of equations from which the variance and autocorrelation function of the process can be obtained. Suppose Z_t follows the INGARCH(p,q) model in Eq. 6 with $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 0$. From Th. 1 part (iii) in Weiß (2009), the autocovariances $\gamma_Z(k) = Cov[Z_t, Z_{t-k}]$ and $\gamma_\mu(k) = Cov[\mu_t, \mu_{t-k}]$ satisfy the linear equations

$$\begin{aligned}\gamma_Z(k) &= \sum_{i=1}^p \alpha_i \gamma_Z(|k-i|) + \sum_{j=1}^{\min(k-1,q)} \beta_j \gamma_Z(k-j) + \sum_{j=k}^q \beta_j \gamma_\mu(j-k), \quad k \geq 1; \\ \gamma_\mu(k) &= \sum_{i=1}^{\min(k,p)} \alpha_i \gamma_\mu(|k-i|) + \sum_{i=k+1}^p \alpha_i \gamma_Z(i-k) + \sum_{j=1}^q \beta_j \gamma_\mu(|k-j|), \quad k \geq 0. \quad (39)\end{aligned}$$

In order to have an explicit expression for the variance of μ_t and Z_t and for the autocorrelations, we consider two special cases as in Zhu (2012) and Weiß (2009). For a proof of the results in these examples, see Section B.3.

Example 1 (INARCH(1)). Consider the INARCH(1) model

$$\mu_t = \alpha_0 + \alpha_1 Z_{t-1} \quad (40)$$

then the linear equations in Eq. 39, becomes

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

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

$$\gamma_Z(k) = \begin{cases} \alpha_1^{k-1} \gamma_Z(1), & \text{for } k \geq 2 \\ \alpha_1 [\phi^3 E(\tilde{\sigma}_t^2)] + \alpha_1 V(\mu_t), & \text{for } k = 1 \end{cases} \quad (41)$$

$$\gamma_\mu(k) = \begin{cases} \alpha_1^k V(\mu_t), & \text{for } k \geq 1 \\ \alpha_1^2 [\phi^3 E(\tilde{\sigma}_t^2)] + \alpha_1^2 V(\mu_t), & \text{for } k = 0 \end{cases} \quad (42)$$

Therefore, the variance of μ_t is

$$V(\mu_t) = \frac{\alpha_1^2 [\phi^3 E(\tilde{\sigma}_t^2)]}{1 - \alpha_1^2} \quad (43)$$

and the variance of Z_t is

$$V(Z_t) = \frac{\phi^3 E(\tilde{\sigma}_t^2)}{1 - \alpha_1^2} \quad (44)$$

where $\phi = \frac{1}{1-\lambda}$.

Lastly, the autocorrelations are

$$\rho_\mu(k) = \alpha_1^k \quad (45)$$

$$\rho_Z(k) = \alpha_1^k \quad (46)$$

Example 2 (INGARCH(1,1)). Consider the INGARCH(1,1) model

$$\mu_t = \alpha_0 + \alpha_1 Z_{t-1} + \beta_1 \mu_{t-1} \quad (47)$$

From Eq. 39,

$$\gamma_Z(k) = \begin{cases} (\alpha_1 + \beta_1)^{k-1} \gamma_Z(1), & \text{for } k \geq 2 \\ \alpha_1^2 [\phi^3 E(\tilde{\sigma}_t^2)] + (\alpha_1 + \beta_1) V(\mu_t), & \text{for } k = 1 \end{cases} \quad (48)$$

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

$$\gamma_\mu(k) = \begin{cases} (\alpha_1 + \beta_1)^k V(\mu_t), & \text{for } k \geq 1 \\ \alpha_1^2 [\phi^3 E(\tilde{\sigma}_t^2)] + (\alpha_1 + \beta_1)^2 V(\mu_t), & \text{for } k = 0 \end{cases} \quad (49)$$

where the second equation in Eq. 49 is equal to $V(\mu_t)$. From this latter equation, we can derive the expression for $V(\mu_t)$

$$V(\mu_t) = \frac{\alpha_1^2 [\phi^3 E(\tilde{\sigma}_t^2)]}{1 - (\alpha_1 + \beta_1)^2} \quad (50)$$

Combining Eq. 38 and 50, we can derive a close expression for the variance of Z_t :

$$V(Z_t) = \frac{\phi^3 E(\tilde{\sigma}_t^2) [1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2]}{1 - (\alpha_1 + \beta_1)^2} \quad (51)$$

where $\phi = \frac{1}{1-\lambda}$.

The autocorrelations are given by

$$\rho_\mu(k) = (\alpha_1 + \beta_1)^k \quad (52)$$

$$\rho_Z(k) = (\alpha_1 + \beta_1)^{k-1} \frac{\alpha_1 [1 - \beta_1 (\alpha_1 + \beta_1)]}{1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2} \quad (53)$$

4 Bayesian Inference

We propose a Bayesian approach to inference for GPD-INGARCH, which allows the researcher to include extra-sample information through the prior choice and allows us to exploit the stochastic representation of the GPD and the use of latent variables to make more tractable the likelihood function.

4.1 Prior assumption

We assume the following prior distributions. A Dirichlet prior distribution for $\boldsymbol{\varphi} = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$, $\boldsymbol{\varphi} \sim \text{Dir}_{d+1}(c)$, with density:

$$\pi(\boldsymbol{\varphi}) = \frac{\Gamma\left(\sum_{i=0}^d c_i\right)}{\prod_{i=0}^d \Gamma(c_i)} \prod_{i=1}^d \varphi_i^{c_i-1} \left(1 - \sum_{i=1}^d \varphi_i\right)^{(c_0-1)} \quad (54)$$

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 function 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. $\pi(\alpha_0) \propto \mathbb{I}_{\mathbb{R}}(\alpha_0)$. 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 $\phi \sim \mathcal{Ga}^*(a, b, c)$, with density function:

$$\pi(\phi) = \frac{b^a}{\Gamma(a)} (\phi - c)^{(a-1)} e^{-b(\phi-c)} \quad \text{for } \phi > c \quad (55)$$

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

Denote the probability distribution of Z_t with

$$f_t(Z_t = z|\boldsymbol{\theta}) = e^{-\sigma_t^2 - z\lambda} \sum_{s=\underline{s}}^{+\infty} \frac{1}{4} \frac{\sigma_t^4 + \mu_t^2}{s!(s+z)!} \left[\frac{\sigma_t^2 + \mu_t}{2} + (s+z)\lambda \right]^{s+z-1} \left[\frac{\sigma_t^2 - \mu_t}{2} + s\lambda \right]^{s-1} e^{-2\lambda s} \quad (56)$$

with $\underline{s} = \max(0, -z)$. Since the posterior distribution

$$\pi(\boldsymbol{\theta}|Z_{1:T}) \propto \prod_{t=1}^T f_t(Z_t|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (57)$$

is not analytically tractable we apply Markov Chain Monte Carlo (MCMC) for posterior approximation in combination with a data-augmentation approach

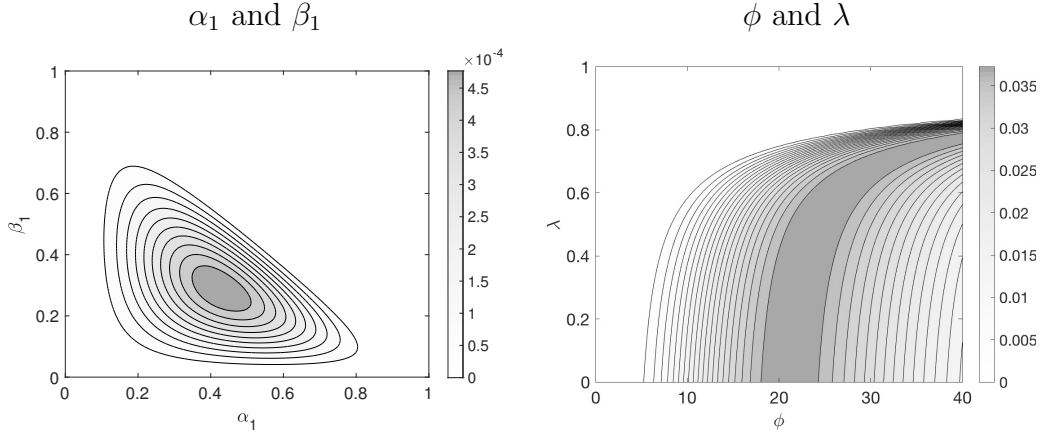


Figure 4: Contour lines of the log-prior density function for α_1 and β_1 (left) and ϕ and λ (right).

(Tanner and Wong, 1987). See Robert and Casella (2013) for an introduction to MCMC. As in Karlis and Ntzoufras (2006), we exploit the stochastic representation in Eq. 8 and introduce two GPD latent variables X_t and Y_t with pmfs

$$f_t(X_t = x | \theta_{1t}, \lambda) = \frac{\theta_{1t}(\theta_{1t} + \lambda x)^{x-1}}{x!} e^{-(\theta_{1t} + \lambda x)} \quad (58)$$

$$f_t(Y_t = y | \theta_{2t}, \lambda) = \frac{\theta_{2t}(\theta_{2t} + \lambda y)^{y-1}}{y!} e^{-(\theta_{2t} + \lambda y)} \quad (59)$$

Let $Z_{1:T} = (Z_1, \dots, Z_T)$, $X_{1:T} = (X_1, \dots, X_T)$ and $Y_{1:T} = (Y_1, \dots, Y_T)$. The complete-data likelihood becomes

$$\begin{aligned} f(Z_{1:T}, X_{1:T}, Y_{1:T} | \boldsymbol{\theta}) &= \prod_{t=1}^T f(Z_t | X_t, Y_t, \boldsymbol{\theta}) f_t(X_t, Y_t | \boldsymbol{\theta}) \\ &= \prod_{t=1}^T \delta(Z_t - X_t + Y_t) f_t(X_t | \boldsymbol{\theta}) f_t(X_t - Z_t | \boldsymbol{\theta}). \end{aligned} \quad (60)$$

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

$$\pi(X_{1:T}, Y_{1:T}, \boldsymbol{\theta} | Z_{1:T}) \propto f(Z_{1:T}, X_{1:T}, Y_{1:T} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \quad (61)$$

4.3 Posterior approximation

We apply a Gibbs algorithm (Robert and Casella, 2013, Ch. 10) with a Metropolis-Hastings (MH) steps. In the sampler, we draw the latent variables and the

parameters of the model by iterating the following steps:

1. draw (X_t, Y_t) from $f(X_t, Y_t|Z_{1:T}, \boldsymbol{\theta})$;
2. draw $\boldsymbol{\varphi}$ from $\pi(\boldsymbol{\varphi}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\boldsymbol{\varphi}})$;
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 η .

The full conditional for the latent variables is

$$(X_t, Y_t) \sim f(Z_t|X_t, Y_t, \boldsymbol{\theta})f(X_t, Y_t|Z_{1:T}, \boldsymbol{\theta}). \quad (62)$$

We draw from the full conditional distribution by MH. Differently from Karlis and Ntzoufras (2006), we use a mixture proposal distribution which allows for a better mixing of the MCMC chain. At the j -th iteration, we generate a candidate X_t^* from $GP(\theta_{1t}, \lambda)$ with probability ν and $(X_t^* - Z_t)$ from $GP(\theta_{2t}, \lambda)$ with probability $1 - \nu$, and accept with probability

$$\varrho = \min \left\{ 1, \frac{f_t(X_t^*|\theta_{1t}, \lambda)f_t(X_t^* - Z_t|\theta_{2t}, \lambda)}{f_t(X_t^{(j-1)}|\theta_{1t}, \lambda)f_t(X_t^{(j-1)} - Z_t|\theta_{2t}, \lambda)} \frac{q(X_t^{(j-1)})}{q(X_t^*)} \right\} \quad (63)$$

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 . The method extends to the GPD the technique proposed in Karlis and Ntzoufras (2006) for the Poisson differences.

As regards to the parameter $\boldsymbol{\varphi}$, its full conditional distribution is

$$\boldsymbol{\varphi} \sim \pi(\boldsymbol{\varphi}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\boldsymbol{\varphi}}) \propto \pi(\boldsymbol{\varphi}) \prod_{t=1}^T f_t(X_t, Y_t|\boldsymbol{\theta}). \quad (64)$$

We consider a MH with Dirichlet independent proposal distribution

$$\boldsymbol{\varphi}^* \sim Dir(\mathbf{c}^*) \quad (65)$$

where $\mathbf{c}^* = (c_0^*, c_1^*, c_2^*)$ and acceptance probability

$$\varrho = 1 \wedge \frac{\pi(\boldsymbol{\varphi}^*|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\boldsymbol{\varphi}})}{\pi(\boldsymbol{\varphi}^{j-1}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\boldsymbol{\varphi}})}. \quad (66)$$

The full conditional distribution of ϕ is

$$\pi(\phi|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi}) \propto \pi(\phi) \prod_{t=1}^T f_t(X_t, Y_t|\boldsymbol{\theta}). \quad (67)$$

We consider the change of variable $\zeta = \log(\phi - c)$ with Jacobian $\exp(\zeta)$ and a MH step with a random walk proposal

$$\zeta^* \sim N(\zeta_{j-1}, \gamma^2) \quad (68)$$

where $\zeta_{j-1} = \log(\phi_{j-1} - c)$, ϕ_{j-1} is the previous iteration value of the parameter and $c = \frac{1}{(1-\lambda)^2}$. The acceptance probability is

$$\varrho = \min \left\{ 1, \frac{\pi(\phi^* | Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi}) \exp(\zeta^*)}{\pi(\phi_{j-1} | Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi}) \exp(\zeta_{j-1})} \right\} \quad (69)$$

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

The full conditional distribution of λ is

$$\pi(\lambda | Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda}) \propto \pi(\lambda) \prod_{t=1}^T f_t(X_t, Y_t | \boldsymbol{\theta}). \quad (70)$$

We consider a MH step with Beta random walk proposal

$$\lambda^* \sim Be(s\lambda^{(j-1)}, s(1 - \lambda^{(j-1)})) \quad (71)$$

where s is a precision parameter. The acceptance probability is:

$$\varrho = \min \left\{ 1, \frac{\pi(\lambda^* | Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda}) Be(s\lambda^*, s(1 - \lambda^*))}{\pi(\lambda^{(j-1)} | Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda}) Be(s\lambda^{(j-1)}, s(1 - \lambda^{(j-1)}))} \right\}. \quad (72)$$

5 Simulation study

The purpose of our simulation exercises is to study the efficiency of the MCMC algorithm presented in Section 4. We evaluated the Geweke (1992) convergence diagnostic measure (CD), the inefficiency factor (INEFF)¹ and the Effective Sample Size (ESS).

We simulated 50 independent data-series of 400 observations each. We run the Gibbs sampler for 1,010,000 iterations on each dataset, discard the first 10,000 draws

¹The 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 and are computed to measure how well the MCMC chain mixes. An INEFF equal to n tells us that we need to draw MCMC samples n times as many as uncorrelated samples.

	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 (0.11)	-28.69 (0.14)	0.78 (0.10)	0.93 (0.04)	-6.27 (0.06)	2.40 (0.05)
CD_{AT}	5.72 (0.23)	-13.18 (0.23)	0.2 (0.23)	0.74 (0.13)	-3.84 (0.15)	1.17 (0.11)

Table 1: Autocorrelation function (ACF), effective sample size (ESS) and inefficiency factor (INEFF) of the posterior MCMC samples for the two settings: low persistence and high persistence. The results are averages over a set of 50 independent MCMC experiments on 50 independent datasets of 400 observations each. We ran the proposed MCMC algorithm for 1,010,000 iterations and evaluate 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. In parenthesis the p-values of the Geweke’s convergence diagnostic.

to remove dependence on initial conditions, and finally apply a thinning procedure with a factor of 250, to reduce the dependence between consecutive draws.

As commonly used in the GARCH and stochastic volatility literature (e.g., see Chib et al., 2002; Casarin et al., 2009; Billio et al., 2016; Bormetti et al., 2019, 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 is effective in reducing the autocorrelation levels and in increasing 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 after

Parameters	Mean	Std	CI
Model \mathcal{M}_1 : GPD-INGARCH(1,1)			
α	0.3920	0.0246	(0.3347, 0.4297)
β	0.4753	0.0096	(0.4582, 0.4999)
λ	0.5892	0.0246	(0.53833, 0.6349)
ϕ	179.7905	22.8040	(138.2406, 226.99)
Model \mathcal{M}_2 : PD-INGARCH(1,1) and $\lambda = 0$			
α	0.1121	0.0095	(0.1004, 0.1340)
β	0.1798	0.0101	(0.1549, 0.1989)
λ	-	-	-
ϕ	94.9340	8.6488	(77.0653, 110.6276)
Model \mathcal{M}_3 : GPD-INARCH(1,0)			
α	0.2286	0.0485	(0.1407, 0.3287)
β	-	-	-
λ	0.5682	0.0243	(0.5195, 0.6166)
ϕ	218.6333	36.2307	(155.7151, 297.2252)
Model \mathcal{M}_4 : PD-INARCH(1,0) and $\lambda = 0$			
α	0.1013	0.0013	(0.1000, 0.1050)
β	-	-	-
λ	-	-	-
ϕ	104.4131	7.4362	(86.4588, 115.8723)

Table 2: Posterior mean (Mean), 95% credible intervals (CI), and standard deviation (Std) for different specifications (different panels) of the GPD-INGARCH.

thinning generally improved. On average, the inefficiency measures (19.05), the p-values of the CD statistics (0.18) and the acceptance rates (0.35) achieved the values recommended in the literature (e.g., see Roberts et al., 1997).

6 Real data examples

6.1 Accident data

Data in this application are the number of accidents near Schiphol airport in The Netherlands during 2001 (Fig. 5). They have been previously considered in Brijs et al. (2008) and Andersson and Karlis (2014). The time series of accident counts is non-stationary and should be differentiated (Kim and Park, 2008). We applied our Bayesian estimation procedure, as described in Section 4.

In Fig. 6 are presented the histograms for the Gibbs draws for each parameters. 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). In the

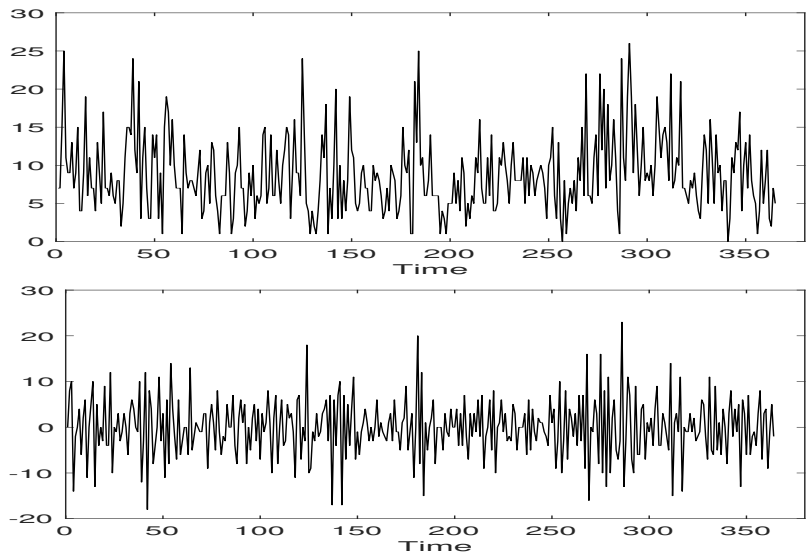


Figure 5: Frequency (top) and month-on-month changes (bottom) of the accidents at the Schiphol airport in The Netherlands in 2001.

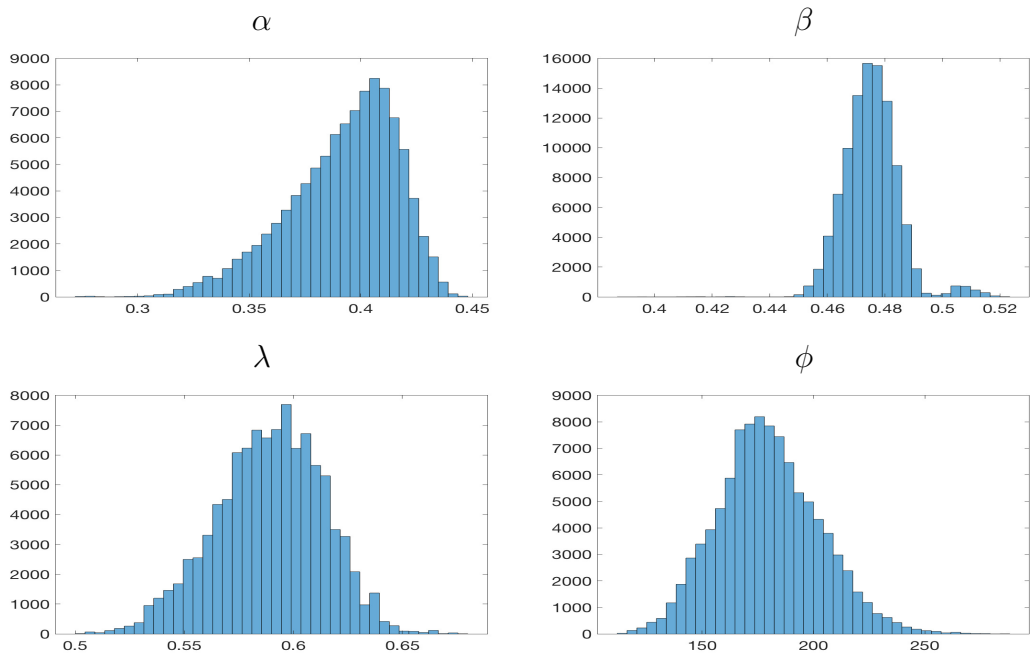


Figure 6: Histograms of the MCMC draws for the parameters of the Schiphol's accident data of Fig. 5.

data, we found evidence of high persistence in the expected accident arrivals, i.e. $\hat{\alpha} + \hat{\beta} = 0.8673$ and heteroskedastic effects, i.e. $\hat{\beta} = 0.4753$. Also, there is evidence in favour of overdispersion, $\hat{\lambda} = 0.5892$ and overdispersion persistence $\hat{\phi} = 179.7905$. We study the contribution of the heteroskedasticity and persistence by testing some restrictions of the INGARCH(1,1) (models from \mathcal{M}_2 to \mathcal{M}_4 in Tab. 2).

Bayesian inference compares models via the so-called Bayes factor, which is the ratio of normalizing constants of the posterior distributions of two different models (see Cameron et al. (2014) for a review). MCMC methods allows for generating samples from the posterior distributions which can be used to estimate the ratio of normalizing constants.

In this paper we use the method proposed by Geyer (1994). The method consists in deriving the normalizing constants by reverse logistic regression. The idea behind this method is to consider the different estimates as if they were sampled from a mixture of two distributions with probability

$$p_j(x, \eta) = \frac{h_j(x) \exp(\eta_j)}{h_1(x) \exp(\eta_1) + h_2(x) \exp(\eta_2)}, \quad j = 1, 2 \quad (73)$$

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

$$\ell_n(\kappa) = \sum_{i=1}^n \log p_1(X_{i1}, \eta_1) + \sum_{i=1}^n \log p_2(X_{i2}, \eta_2) \quad (74)$$

where n is the number of MCMC draws for each model and $X_{ij} = \log f(Z_{1:T}, X_{1:T}^{(i)}, Y_{1:T}^{(i)} | \boldsymbol{\theta}^{(i)})$ is the log-likelihood evaluated at the i -th MCMC sample for each model of Tab. 2.

We performed six reverse logistic regressions, in which we compare pairwise our models. The approximated logarithmic Bayes factors $BF(\mathcal{M}_i, \mathcal{M}_j)$ are given in Tab. 3. It is possible to see that our GPD-INGARCH(1, 1), \mathcal{M}_1 , is preferable with respect to the other models. Notice that \mathcal{M}_2 corresponds to an INGARCH(1, 1) where the observations are form a standard Poisson-difference model PD-INGARCH(1, 1), \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).

6.2 Cyber threats 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), in this paper we use the classification

$\text{BF}(\mathcal{M}_i, \mathcal{M}_j)$	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4
\mathcal{M}_1	333.45 (5.818)	25.19 (0.253)	121.44 (0.521)
\mathcal{M}_2		-226.86 (2.024)	-300.96 (2.522)
\mathcal{M}_3			-73.25 (0.358)

Table 3: Logarithmic Bayes Factor, $\text{BF}(\mathcal{M}_i, \mathcal{M}_j)$, 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 $\lambda = 0$. Number in parenthesis are standard deviations of the estimated Bayes factors.

provided in the Hackmageddon dataset. Hackmageddon is a well-known cyber-incident website that collects public reports and provides the number of cyber incidents for different categories of threats: crimes, espionage and warfare.

Figure 7 shows the total and category-specific number of cyber attacks at a daily frequency from January 2017 to December 2018. Albeit limited in the 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 threats frequencies are between 0 and 12 which motivates the use of a discrete distribution. We remove the upward trend by considering the first difference and fit the GPD-INGARCH model proposed in Section 2.

We applied our estimation procedure, as described in Section 4. As in the previous application, we fix $\alpha_0 = 1.05$ that is coherent with the conditional mean of the time series. We ran the Gibbs sampler for 110000 iterations, where we discarded the first 10000 iterations as burn-in sample. In Fig. 8 are presented the histograms for the Gibbs draws for each parameters.

Figure 8 shows that, as before, it is reasonable to fit a GPD-INGARCH process to the difference of cyber attacks since both the autoregressive parameter α and β , that represent the heteroskedastic feature of the data, are different from zero. Additionally, the value of λ suggest the presence of over-dispersion in the data.

Given the importance of forecasting cyber-attacks, in this section we present the results of one-step-ahead forecasting exercise over a period of 120. We follow an approach based on predictive distributions which quantifies all uncertainty associated with the future number of attacks and is used in a wide range of applications (see, e.g. McCabe and Martin, 2005; McCabe et al., 2011, and references therein). We account for parameter uncertainty and approximate the predictive distribution by MCMC. At the j -th MCMC iteration we draw $Z_{T+h}^{(j)}$ from the conditional distribution given past observations and the parameter draw $\theta^{(j)}$

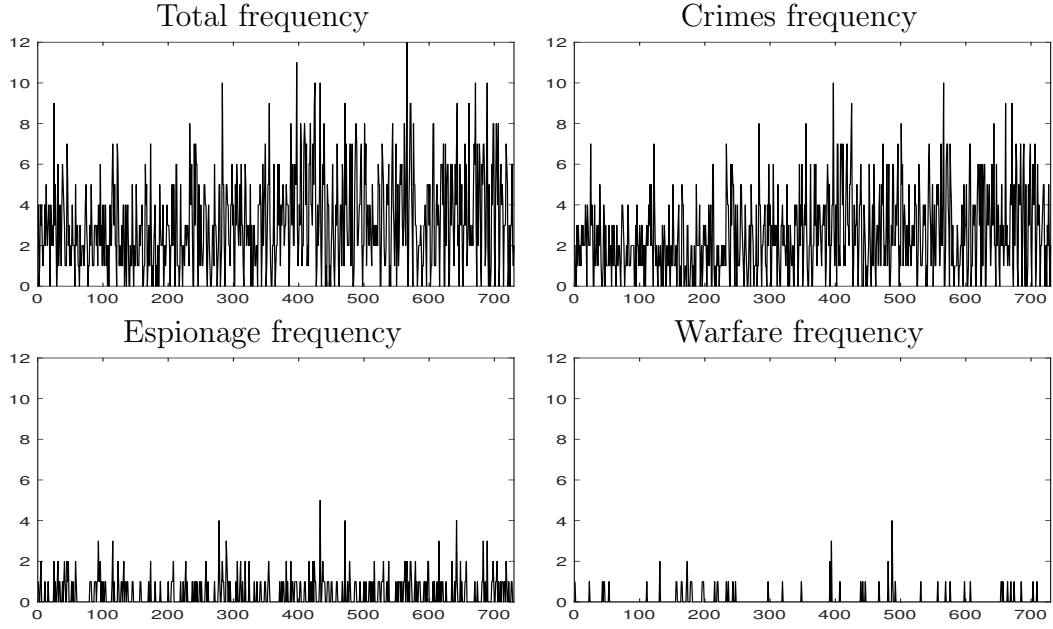


Figure 7: Daily cyber-threats counts between 1st January 2017 and 31st December 2018.

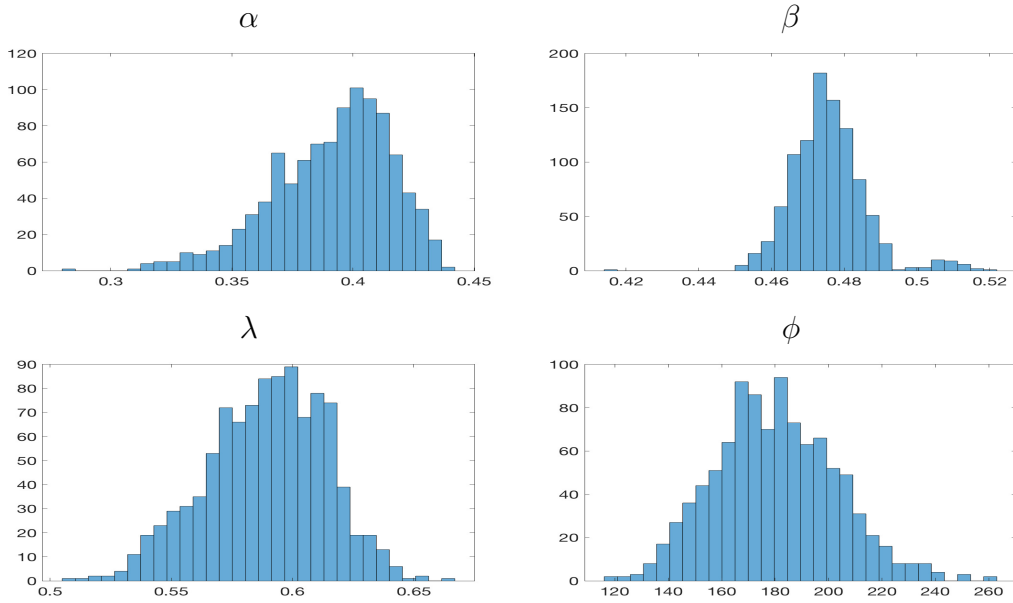


Figure 8: Histograms of the MCMC draws for the parameters.

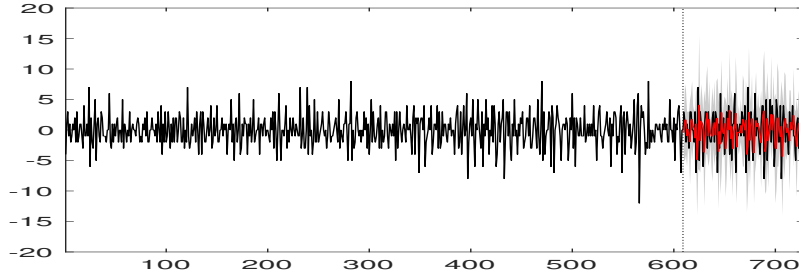


Figure 9: Changes in the number of cyber threats and their sequential one-step-ahead forecast (red line) and 95% HPD region (gray area) between 1st November 2018 (vertical dashed line) and 31st December 2018.

$$Z_{T+h}^{(j)} | \mathcal{F}_T, \boldsymbol{\theta}^{(j)} \sim GPD \left(\mu_{T+h}^{*(j)}, \sigma_{T+h}^{2*(j)}, \lambda^{(j)} \right) \quad (75)$$

where $j = 1, \dots, J$, denotes the MCMC draw, $h = 1, \dots, H$ the forecasting horizon and

$$\mu_{T+h}^{(j)} = \begin{cases} \alpha_0^{(j)} + \alpha_1^{(j)} Z_T + \beta_1^{(j)} \mu_T^{(j)}, & \text{for } h = 1 \\ \alpha_0^{(j)} + \alpha_1^{(j)} Z_{T+h-1}^{(j)} + \beta_1^{(j)} \mu_{T+h-1}^{(j)}, & \text{for } h = 2, \dots, H \end{cases} \quad (76)$$

$$\sigma_{T+h}^{2(j)} = |\mu_{T+h}^{(j)}| \phi^{(j)} \quad (77)$$

7 Conclusions

We introduce a new family of stochastic processes with values in the set of integers with sign. The increments of the process follow a generalized Poisson difference distribution with time-varying parameters. We assume a GARCH-type dynamics, provide a thinning representation and study the properties of the process. We provide a Bayesian inference procedure and an efficient Monte Carlo Markov Chain sampler for posterior approximation. Inference and forecasting exercises on accidents and cyber-threats data show that the proposed GPD-INGARCH model is well suited for capturing persistence in the conditional moments and in the over-dispersion feature of the data.

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation.

- Agrafiotis, I., Nurse, J. R. C., Goldsmith, M., Creese, S., and Upton, D. (2018). A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate. *Journal of Cybersecurity*, 4(1).
- Al-Osh, M. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis*, 8(3):261–275.
- Al-Osh, M. A. and 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.
- 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. and Al-Osh, M. (1993). Generalized Poisson ARMA processes. *Annals of the Institute of Statistical Mathematics*, 45(2):223–232.
- Alzaid, A. A. and 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. and Moore, T. (2006). The economics of information security. *Science*, 314(5799):610–613.
- Andersson, J. and Karlis, D. (2014). A parametric time series model with covariates for integers in Z . *Statistical Modelling*, 14(2):135–156.
- Billio, M., Casarin, R., and Osuntuyi, A. (2016). Efficient Gibbs sampling for Markov switching GARCH models. *Computational Statistics and Data Analysis*, 100:37 – 57.
- Bormetti, G., Casarin, R., Corsi, F., and Livieri, G. (2019). A stochastic volatility model with realized measures for option pricing. *Journal of Business & Economic Statistics*, 0(0):1–31.
- Brenner, S. W. (2004). Cybercrime metrics: Old wine, new bottles? *Virginia Journal of Law and Technology*, 9(13):1–53.
- Brijs, T., Karlis, D., and Wets, G. (2008). Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accident Analysis & Prevention*, 40(3).
- Brockwell, P. J., Davis, R. A., and Fienberg, S. E. (1991). *Time Series: Theory and Methods: Theory and Methods*. Springer Science & Business Media.

- Cameron, E., Pettitt, A., et al. (2014). Recursive pathways to marginal likelihood estimation with prior-sensitivity analysis. *Statistical Science*, 29(3).
- Cardinal, M., Roy, R., and 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.
- Casarin, R., Marin, J.-M., et al. (2009). Online data processing: Comparison of Bayesian regularized particle filters. *Electronic Journal of Statistics*, 3:239–258.
- Chen, C. W. and 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., and 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., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316.
- Consul, P. (1990). On some properties and applications of quasi-binomial distribution. *Communications in Statistics-Theory and Methods*, 19(2):477–504.
- Consul, P. and Famoye, F. (1986). On the unimodality of generalized Poisson distribution. *Statistica neerlandica*, 40(2):117–122.
- Consul, P. and Famoye, F. (1992). Generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 21(1):89–109.
- Consul, P. and Mittal, S. (1975). A new urn model with predetermined strategy. *Biometrische Zeitschrift*, 17(2):67–75.
- Consul, P. and Shenton, L. (1975). On the probabilistic structure and properties of discrete Lagrangian distributions. In *A modern course on statistical distributions in scientific work*, pages 41–57.
- Consul, P. C. (1986). On the differences of two generalized Poisson variates. *Communications in Statistics - Simulation and Computation*, 15(3):761–767.
- Consul, P. C. (1989). *Generalized Poisson Distributions*. Dekker New York.
- Consul, P. C. and Famoye, F. (2006). *Lagrangian probability distributions*. Springer.
- Consul, P. C. and Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics*, 15(4):791–799.

- Consul, P. C. and Shenton, L. (1973). Some interesting properties of Lagrangian distributions. *Communications in Statistics-Theory and Methods*, 2(3):263–272.
- Cunha, E. T. d., Vasconcellos, K. L., and 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., and Wang, Y. (1999). Modeling time series of count data. *Statistics Textbooks and Monographs*, 158:63–114.
- Demirtas, H. (2017). On accurate and precise generation of generalized Poisson variates. *Communications in Statistics-Simulation and Computation*, 46(1):489–499.
- Edwards, B., Hofmeyr, S. A., and Forrest, S. (2015). Hype and heavy tails: A closer look at data breaches. *J. 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. (1993). Restricted generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 22(5):1335–1354.
- Famoye, F. (1997). Generalized Poisson random variate generation. *American Journal of Mathematical and Management Sciences*, 17(3-4):219–237.
- Famoye, F. (2015). A multivariate generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 44(3):497–511.
- Famoye, F. and Consul, P. (1995). Bivariate generalized Poisson distribution with some applications. *Metrika*, 42(1).
- Famoye, F., Wulu, J. T., and Singh, K. P. (2004). On the generalized Poisson regression model with an application to accident data. *Journal of Data Science*, 2(2004):287–295.
- Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, 27(6):923–942.

- Francq, C. and Zakoian, J.-M. (2019). *GARCH models: structure, statistical inference and financial applications*. Wiley.
- Freeland, R. and McCabe, B. P. (2004). Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, 25(5):701–722.
- Freeland, R. K. (1998). *Statistical analysis of discrete time series with application to the analysis of workers' compensation claims data*. PhD thesis, University of British Columbia.
- Freeland, R. K. (2010). True integer value time series. *AStA Advances in Statistical Analysis*, 94(3):217–229.
- FSB (2018). Cyber lexicon. Available at <https://www.fsb.org/wp-content/uploads/P121118-1.pdf>.
- Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, Oxford.
- Geyer, C. J. (1994). Estimating normalizing constants and reweighting mixtures. *Technical Report 568, School of Statistics, Univ. Minnesota*.
- Hassanien, A. E., Fouad, M. M., Manaf, A. A., Zamani, M., Ahmad, R., and Kacprzyk, J. (2016). *Multimedia Forensics and Security: Foundations, Innovations, and Applications*, volume 115. Springer.
- Heinen, A. (2003). Modelling time series count data: An autoregressive conditional Poisson model. *Available at SSRN 1117187*.
- Hubert Jr, P. C., Lauretto, M. S., and Stern, J. M. (2009). Fbst for generalized Poisson distribution. In *AIP Conference Proceedings*, volume 1193, pages 210–217. AIP.
- Husák, M., Komárková, J., Bou-Harb, E., and Čeleda, P. (2018). Survey of attack projection, prediction, and forecasting in cyber security. *IEEE Communications Surveys & Tutorials*, 21(1):640–660.
- Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution. *Journal of the Royal Statistical Society*, 100(3):415–416.
- Jin-Guan, D. and Yuan, L. (1991). The integer-valued autoregressive (INAR (p)) model. *Journal of Time Series Analysis*, 12(2):129–142.

- Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 33(3):664–677.
- Karlis, D. and Ntzoufras, I. (2006). Bayesian analysis of the differences of count data. *Statistics in Medicine*, 25(11):1885–1905.
- Kedem, B. and Fokianos, K. (2005). *Regression models for time series analysis*, volume 488. John Wiley & Sons.
- Kim, H.-Y. and Park, Y. (2008). A non-stationary integer-valued autoregressive model. *Statistical Papers*, 49(3):485.
- Koopman, S. J., Lit, R., and Lucas, A. (2014). The dynamic Skellam model with applications. WorkingPaper 14-032/IV/DSF73, Tinbergen Institute.
- 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., and Pohlmeier, W. (2006). Modelling financial transaction price movements: A dynamic integer count data model. *Empirical Economics*, 30(4):795–825.
- MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*, volume 110. CRC Press.
- McCabe, B. and 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., and 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.
- McKenzie, E. (2003). Ch. 16. discrete variate time series. *Handbook of statistics*, 21:573–606.
- Passeri, P. (2019). Hackmageddon - information security timelines and statistics. <https://www.hackmageddon.com/>.

- Pedeli, X. and Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical modelling*, 11(4):325–349.
- Robert, C. and 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. *The Annals of Applied Probability*, 7(1):110–120.
- Rydberg, T. and Shephard, N. (2000). Bin models for trade-by-trade data. *Modelling the number of trades in fixed interval of time. Paper*, 740:28.
- Rydberg, T. H. and 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., and Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: A review. *Statistical Modelling*, 15(6):590–618.
- Shahtahmassebi, G. and 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. and Moyeed, R. (2016). An application of the generalized Poisson difference distribution to the bayesian modelling of football scores. *Statistica Neerlandica*, 70(3):260–273.
- Skellam, J. (1946). The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society: Series A*, 109(Pt 3):296.
- Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability*, 7(5):893–899.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.
- Tripathi, R. C., Gupta, P. L., and Gupta, R. C. (1986). Incomplete moments of modified power series distributions with applications. *Communications in Statistics-Theory and Methods*, 15(3):999–1015.

- Wang, W. and Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression. *Journal of Population Economics*, 10(3):273–283.
- Wei, C. H. (2008). Thinning operations for modeling time series of counts? A survey. *Advances in Statistical Analysis*, 92(3):319.
- Wei, C. H. (2009). Modelling time series of counts with overdispersion. *Statistical Methods and Applications*, 18(4):507–519.
- Werner, G., Yang, S., and McConky, K. (2017). Time series forecasting of cyber attack intensity. In *Proceedings of the 12th Annual Conference on cyber and information security research*, page 18. ACM.
- Xu, M., Hua, L., and Xu, S. (2017). A vine copula model for predicting the effectiveness of cyber defense early-warning. *Technometrics*, 59(4):508–520.
- Zamani, H., Faroughi, P., and Ismail, N. (2016). Bivariate generalized Poisson regression model: Applications on health care data. *Empirical Economics*, 51(4):1607–1621.
- Zamani, H. and Ismail, N. (2012). Functional form for the generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 41(20):3666–3675.
- Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, 75(4):621–629.
- Zhang, H., Wang, D., and Zhu, F. (2010). Inference for INAR(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference*, 140(3):667–683.
- Zheng, H., Basawa, I. V., and Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*, 137(1):212 – 229.
- 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. and Li, Q. (2009). Moment and Bayesian estimation of parameters in the INGARCH(1, 1) model. *Journal of Jilin University*, 47:899–902.

A Distributions used in this paper

A.1 Poisson Difference distribution

The Poisson difference distribution, a.k.a. as Skellam distribution, is a discrete distribution defined as the difference of two independent Poisson random variables $N_1 - N_2$, with parameters λ_1 and λ_2 . It has been introduced by Irwin (1937) and Skellam (1946).

The probability mass function of the Skellam distribution for the difference $X = N_1 - N_2$ is

$$P(X = x) = e^{-(\lambda_1 + \lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{x/2} I_{|x|}(2\sqrt{\lambda_1 \lambda_2}), \quad \text{with } X \in \mathbb{Z} \quad (\text{A.1})$$

where $\mathbb{Z} = \dots, -1, 0, 1, \dots$ is the set of positive and negative integer numbers, and $I_k(z)$ is the modified Bessel function of the first kind, defined as (Abramowitz and Stegun, 1965)

$$I_\nu(z) = \left(\frac{z}{2}\right)^2 \sum_{k=0}^{\infty} \frac{\left(\frac{z^2}{4}\right)^k}{k! \Gamma(\nu + k + 1)} \quad (\text{A.2})$$

It can be used, for example, to model the difference in number of events, like accidents, between two different cities or years. Moreover, can be used to model the point spread between different teams in sports, where all scored points are independent and equal, meaning they are single units. Another applications can be found in graphics since it can be used for describing the statistics of the difference between two images with a simple Shot noise, usually modelled as a Poisson process.

The distribution has the following properties:

- Parameters: $\lambda_1 \geq 0, \lambda_2 \geq 0$
- Support: $\{-\infty, +\infty\}$
- Moment-generating function: $e^{-(\lambda_1 + \lambda_2) + \lambda_1 e^t + \lambda_2 e^{-t}}$
- Probability generating function: $e^{-(\lambda_1 + \lambda_2) + \lambda_1 t + \lambda_2 / t}$
- Characteristic function: $e^{-(\lambda_1 + \lambda_2) + \lambda_1 e^{it} + \lambda_2 e^{-it}}$
- Moments
 1. Mean: $\lambda_1 - \lambda_2$
 2. Variance: $\lambda_1 + \lambda_2$
 3. Skewness: $\frac{\lambda_1 - \lambda_2}{(\lambda_1 + \lambda_2)^{3/2}}$

4. Excess Kurtosis: $\frac{1}{\lambda_1 + \lambda_2}$

- The Skellam probability mass function is normalized: $\sum_{k=-\infty}^{+\infty} p(k; \lambda_1, \lambda_2) = 1$

A.2 Generalized Poisson distribution

The Generalized Poisson distribution (GP) has been introduced by Consul and Jain (1973) in order to overcome the equality of mean and variance that characterizes the Poisson distribution. In some cases the occurrence of an event, in a population that should be Poissonian, changes with time or dependently on previous occurrences. Therefore, mean and variances are unequal in the data. In different fields a vastness of mixture and compound distribution have been considered, Consul and Jain introduced the GP distribution in order to obtain a unique distribution to be used in the cases said above, by allowing the introduction of an additional parameter.

See Consul and Famoye (2006) for some applications of the Generalized Poisson distribution. Application of the GP distribution can be find as well in economics and finance. Consul (1989) showed that the number of unit of different commodities purchased by consumers in a fixed period of time follows a Generalized Poisson distribution. He gave interpretation of both parameters of the distribution: θ denote the basic sales potential for the commodity, while λ the average rates of liking generated by the product between consumers. Tripathi et al. (1986) provide an application of the GP distribution in textile manufacturing industry. In particular, given the established use of the Poisson distribution in the field, they compare the Poisson and the GP distributions when firms want to increase their profit. They found that the Generalized Poisson, considering different values of the parameters, always yield larger profits. Moreover, the Generalized Poisson distribution, as studied by Consul (1989), can be used to describe accidents of various kinds, such as: shunting accidents, home injuries and strikes in industries. Another application to accidents has been carried out by Famoye and Consul (1995), where they introduced a bivariate extension to the GP distribution and studied two different estimation methods, i.e. method of moments and MLE, and the goodness of fit of the distribution in accidents statistics. Hubert Jr et al. (2009) test for the value of the GP distribution extra parameter by means of a Bayesian hypotheses test procedure, namely the Full Bayesian Significance Test. Famoye (1997) and Demirtas (2017) provided different methods of sampling from the Generalized Poisson distribution and algorithms for sampling. As regard processes, the GP distribution has been used in different models. For example, Consul and Famoye (1992) introduced the GP regression model, while Famoye (1993) studied the restricted generalized Poisson regression. Wang and Famoye (1997) applied the GP regression model to households' fertility decisions and Famoye et al. (2004) carried out an application of the GP regression model to accident data. Zamani and Ismail (2012) develop a

functional form of the GP regression model, Zamani et al. (2016) introduced a few forms of bivariate GP regression model and different applications using dataset on healthcare, in particular the Australian health survey and the US National Medical Expenditure survey. Famoye (2015) provide a multivariate GP regression model, based on a multivariate version of the GP distribution, and two applications: to the healthcare utilizations and to the number of sexual partners.

The Generalized Poisson distribution of a random variable X with parameters θ and λ is given by

$$P_x(\theta, \lambda) = \begin{cases} \frac{\theta(\theta+\lambda x)^{x-1}}{x!} e^{(-\theta-\lambda x)}, & x = 0, 1, 2, \dots \\ 0, & \text{for } x > m \text{ if } \lambda < 0. \end{cases} \quad (\text{A.3})$$

The GP is part of the class of general Lagrangian distributions. The GP has Generating functions and moments

- Parameters:
 1. $\theta > 0$
 2. $\max(-1, -\theta/m) \leq \lambda \leq 1$
 3. $m(\geq 4)$ is the largest positive integer for which $\theta + m\lambda > 0$ when $\lambda < 0$
- Moment generating function (mgf): $M_x(\beta) = e^{\theta(e^s-1)}$, where $z = e^s$ and $u = e^\beta$
- Probability generating function (pgf): $G(u) = e^{\theta(z-1)}$, where $z = ue^{\lambda(z-1)}$
- Moments:
 1. Mean: $\mu = \theta(1 - \lambda)^{-1}$
 2. Variance: $\sigma^2 = \theta(1 - \lambda)^{-3}$
 3. Skewness: $\beta_1 = \frac{1+2\lambda}{\sqrt{\theta(1-\lambda)}}$
 4. Kurtosis: $\beta_2 = 3 + \frac{1+8\lambda+6\lambda^2}{\theta(1-\lambda)}$

The pgf of the GP is derived by Consul and Jain (1973) by means of the Lagrange expansion, namely:

$$z = \sum_{x=1}^{\infty} \frac{u^x}{x!} \{D^{x-1}(g(z))^x\}_{z=0} \quad (\text{A.4})$$

$$\begin{aligned} f(z) &= \sum_{x=0}^{\infty} \frac{u^x}{x!} D^{x-1} \{(g(z))^x f'(z)\}_{z=0} \\ &= f(0) + \sum_{x=1}^{\infty} \frac{u^x}{x!} D^{x-1} \{(g(z))^x f'(z)\}_{z=0} \end{aligned} \quad (\text{A.5})$$

where $D^{x-1} = \frac{d^{x-1}}{dz^{x-1}}$. In particular, for the GP distribution we have (Consul and Famoye, 2006) :

$$f(z) = e^{\theta(z-1)} \quad \text{and} \quad g(z) = e^{\lambda(z-1)} \quad (\text{A.6})$$

Now, by setting $G(u) = f(z)$ we have the expression above for the pgf. (see proof in

Properties Consul and Jain (1973), Consul (1989), Consul and Famoye (1986) and Consul and Famoye (2006) derived some interesting properties of the Generalized Poisson distribution.

Theorem 1 (Convolution Property). The sum of two independent random Generalized Poisson variates, $X + Y$, with parameters (θ_1, λ) and (θ_2, λ) is a Generalized Poisson variate with parameters $(\theta_1 + \theta_2, \lambda)$.

For a proof of Th. 1 see Consul and Jain (1973).

Theorem 2 (Unimodality). The GP distribution models are unimodal for all values of θ and λ and the mode is at $x = 0$ if $\theta e^{-\lambda} < 1$ and at the dual points $x = 0$ and $x = 1$ when $\theta e^{-\lambda} = 1$ and for $\theta e^{-\lambda} > 1$ the mode is at some point $x = M$ such that:

$$(\theta - e^{-\lambda})(e^\lambda - 2\lambda)^{-1} < M < a \quad (\text{A.7})$$

where a is the smallest value of M satisfying the inequality

$$\lambda^2 M^2 + M[2\lambda\theta - (\theta + 2\lambda)e^\lambda] > 0 \quad (\text{A.8})$$

For a proof of Th. 2 see Consul and Famoye (1986).

Consul and Shenton (1975) and Consul (1989) derived some recurrence relations between noncentral moments μ'_k and the cumulants K_k :

$$(1 - \lambda)\mu'_{k+1} = \theta\mu'_k + \theta\frac{\partial\mu'_k}{\partial\theta} + \lambda\frac{\partial\mu'_k}{\partial\lambda}, \quad k = 0, 1, 2, \dots \quad (\text{A.9})$$

$$(1 - \lambda)K_{k+1} = \lambda\frac{\partial K_k}{\partial\lambda} + \theta\frac{\partial K_k}{\partial\theta}, \quad k = 1, 2, 3, \dots \quad (\text{A.10})$$

Moreover, a recurrence relation between the central moments of the GP distribution has been derived:

$$\mu_{k+1} = \frac{\theta k}{(1 - \lambda)^3} \mu_{k-1} + \frac{1}{1 - \lambda} \left\{ \frac{d\mu_k(t)}{dt} \right\}_{t=1}, \quad k = 1, 2, 3, \dots \quad (\text{A.11})$$

where $\mu_k(t)$ is the central moment μ_k with θt and λt in place of θ and λ .

A.3 Generalized Poisson Difference distribution

The random variable X follows a Generalized Poisson distribution (GP) $P_x(\theta, \lambda)$ if and only if

$$P_x(\theta, \lambda) = \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-\theta - x\lambda} \quad x = 0, 1, 2, \dots \quad (\text{A.12})$$

with $\theta > 0$ and $0 \leq \lambda < 1$.

Let $X \sim P_x(\theta_1, \lambda)$ and $Y \sim P_y(\theta_2, \lambda)$ be two independent random variables, Consul (1986) showed that the probability distribution of D , $D = X - Y$, is:

$$P(D = d) = e^{-\theta_1 - \theta_2 - d\lambda} \sum_{y=0}^{\infty} \frac{\theta_2(\theta_2 + y\lambda)^{y-1}}{y!} \frac{\theta_1(\theta_1 + (y+d)\lambda)^{y+d-1}}{(y+d)!} e^{-2y\lambda} \quad (\text{A.13})$$

where d takes all integer values in the interval $(-\infty, +\infty)$. As for the GP distribution, we need to set lower limits for λ in order to ensure that there are at least five classes with non-zero probability when λ is negative. Hence, we set

$$\max(-1, -\theta_1/m_1) < \lambda < 1$$

$$\max(-1, -\theta_2/m_2) < \lambda < 1$$

where, $m_1, m_2 \geq 4$ are the largest positive integers such that $\theta_1 + m_1\lambda > 0$ and $\theta_2 + m_2\lambda > 0$.

Proposition 10. *The probability distribution in (A.13) can be written as*

$$P(D = d) = e^{-(\theta_1 + \theta_2)} \cdot e^{-d\lambda} \sum_{y=0}^{\infty} \frac{\theta_1 \theta_2}{y!(y+d)!} (\theta_2 + y\lambda)^{y-1} (\theta_1 + (y+d)\lambda)^{y+d-1} \cdot e^{-2y\lambda} = C_d(\theta_1, \theta_2, \lambda). \quad (\text{A.14})$$

Therefore, equation (A.14) is the pgf of the difference of two GP variates, from which is possible to obtain the following particular cases:

$$\begin{cases} C_d(\theta_1, 0, \lambda) = \frac{\theta_1(\theta_1 + d\lambda)^{d-1}}{d!} e^{-\theta_1 - d\lambda}, & \text{for } d = 0, 1, 2, \dots \\ C_d(0, \theta_2, \lambda) = \frac{\theta_2(\theta_2 - d\lambda)^{-d-1}}{(-d)!} e^{-\theta_2 + d\lambda}, & \text{for } d = 0, -1, -2, \dots \\ C_d(\theta_1, \theta_2, 0) = e^{-\theta_1 - \theta_2} (\theta_1)^{d/2} (\theta_2)^{-d/2} I_d(2\sqrt{\theta_1 \theta_2}), \end{cases} \quad (\text{A.15})$$

where d is any integer (positive, 0 or negative) and $I_d(z)$ is the modified Bessel function of the first kind, of order d and argument z .

The last result in equation (A.15) is the Skellam distribution (Skellam, 1946). Therefore, the Skellam distribution is a particular case of the difference of two GP variates.

By Consul and Shenton (1973):

$$G_1(u) = e^{\theta_1(t_1-1)}, \text{ where } t_1 = ue^{\lambda(t_1-1)}$$

and

$$G_2(u) = e^{\theta_2(t_2-1)}, \text{ where } t_2 = u^{-1}e^{\lambda(t_2-1)}.$$

Therefore, given that $G(u) = G_1(u)G_2(u)$, the probability generating function (pgf) of the random variable $D = X - Y$ is

$$G(u) = \exp[(\theta_1(t_1 - 1) + (\theta_2(t_2 - 1))]. \quad (\text{A.16})$$

From the cumulant generating function

$$\psi(\beta) = \frac{(T_1 - \beta)\theta_1}{\lambda} + \frac{(T_2 + \beta)\theta_2}{\lambda}$$

where $T_1 = \beta + \lambda(e^{T_1} - 1)$ and $T_2 = -\beta + \lambda(e^{T_2} - 1)$, it is possible to define the mean, variance, skewness and kurtosis of the distribution.

$$L_1 = \frac{(\theta_1 - \theta_2)}{1 - \lambda} \text{ is the first cumulant and the mean.} \quad (\text{A.17})$$

$$L_2 = \frac{(\theta_1 + \theta_2)}{(1 - \lambda)^3} \text{ is the second cumulant and the variance.} \quad (\text{A.18})$$

$$\beta_1 = \frac{(\theta_1 - \theta_2)^2 (1 + 2\lambda)^2}{(\theta_1 + \theta_2)^3 (1 - \lambda)} \text{ is the skewness.} \quad (\text{A.19})$$

$$\beta_2 = 3 + \frac{1 + 8\lambda + 6\lambda^2}{(\theta_1 + \theta_2)(1 - \lambda)} \text{ is the kurtosis.} \quad (\text{A.20})$$

In Fig. A.1-A.2, we show the GPD for various setting of λ , σ^2 and μ .

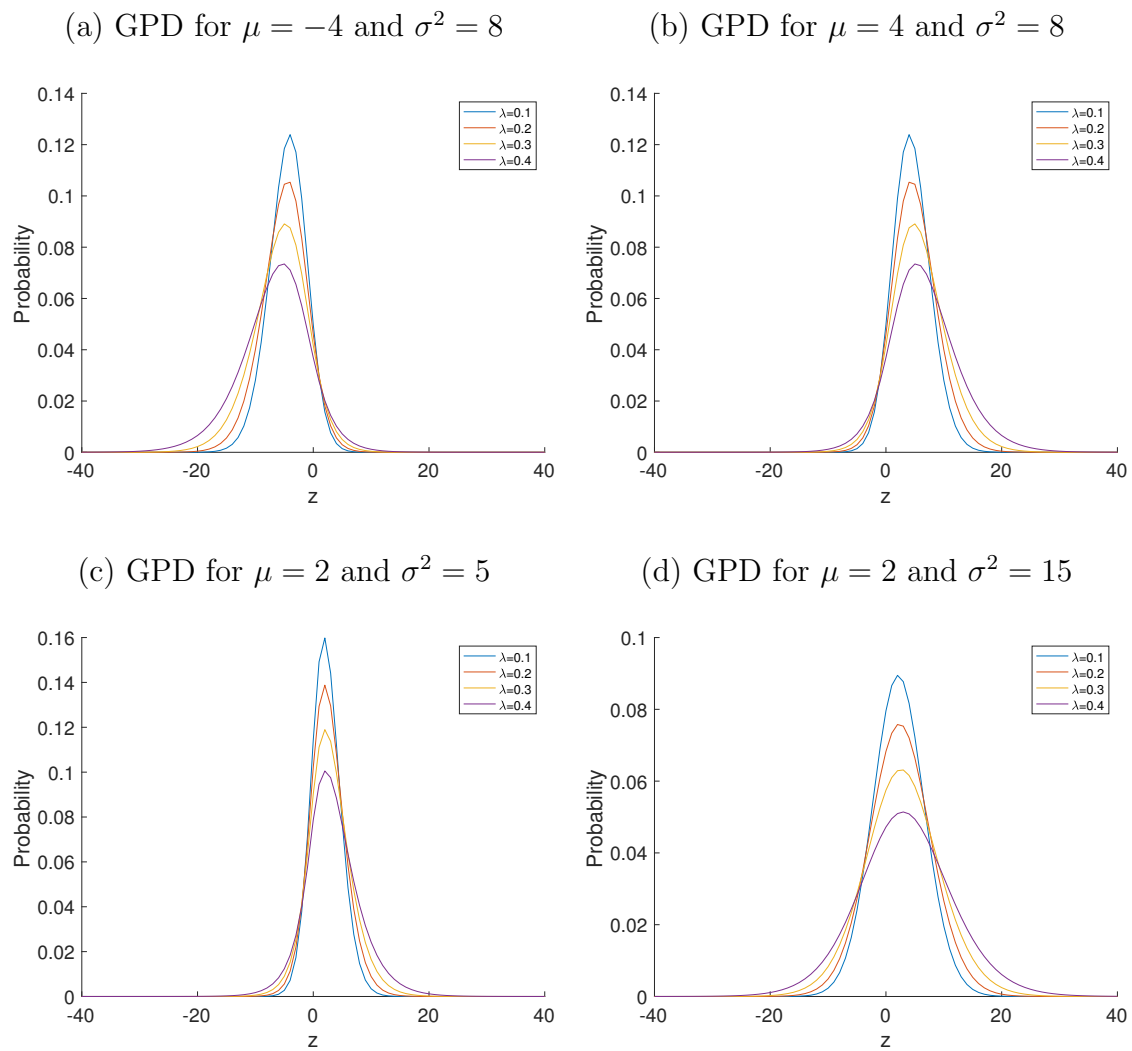


Figure A.1: Generalized Poisson difference distribution for different values of λ . Panels (a) and (b) show the GPD when λ varies, for a fixed value of $\sigma^2 = 8$ and two different values of μ . Panels (c) and (d) show the GPD when λ varies, for a fixed value of $\mu = 2$ and two different value of σ^2 .

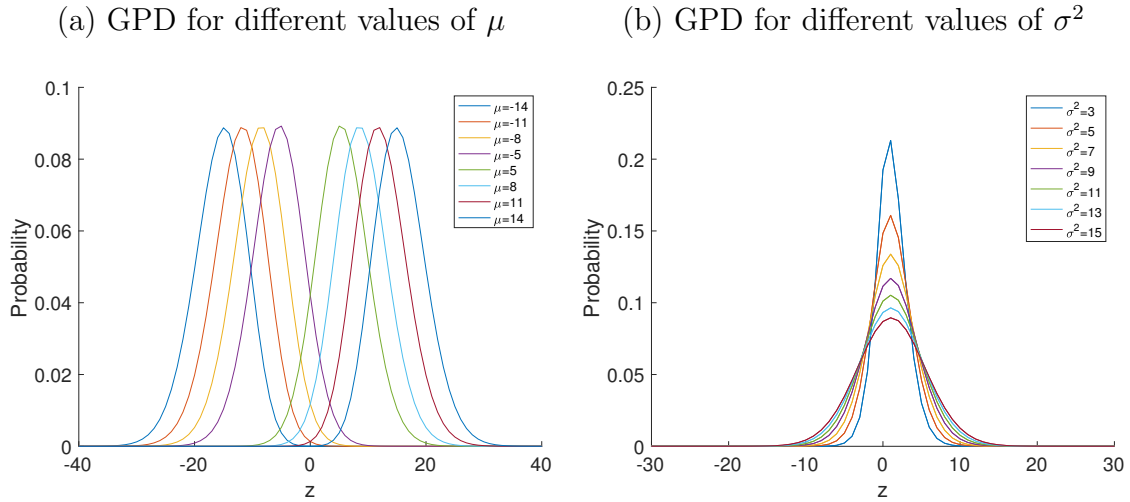


Figure A.2: Generalized poisson difference distribution for different values of μ and σ and fixed λ . In panel (a) μ varies, while $\sigma^2 = 15$ and $\lambda = 0.1$ are fixed. In panel (b) σ^2 varies and $\mu = 1$ and $\lambda = 0.1$ are fixed.

Figure A.1 shows how the GPD distribution varies when λ varies. Given the constraints imposed to the parameters (see section 2) here $\lambda = (0, 0.1, 0.2, 0.3, 0.4)$ since smaller values and, possibly, negative values do not meet the conditions for the selected values of θ_1 and θ_2 . From panel (a) and especially in panel (b) can be seen that when λ increases the distribution becomes longer tailed. From panel (c) and (d) we can see that for fixed values μ , when λ decreases, the GPD is more skewed respectively to the right for $\mu > 0$ ($\theta_1 > \theta_2$) and to the left for $\mu < 0$ ($\theta_1 < \theta_2$). Therefore, the sign of μ determines the skewness of the GPD.

From figure A.2 we can see again, that for positive values of μ the distribution becomes more right-skewed, panels (a) and (b), and more left-skewed for negative values of μ in panels (c) and (d). Moreover, here can be seen better that has θ_1 increases the distribution has longer tails.

A.4 Quasi-Binomial distribution

A first version of the Quasi-Binomial distribution, defined as QB-I by Consul and Famoye (2006), was investigated by Consul (1990) as an urn model. In their definition of the QB thinning, however, Alzaid and Al-Osh (1993) used the QB distribution introduced in the literature by Consul and Mittal (1975) and defined by Consul and Famoye (2006) as QB-II.

$$P(X = x) = \binom{n}{x} \frac{ab}{a+b} \frac{(a+x\theta)^{x-1} (b+n\theta-x\theta)^{n-x-1}}{(a+b+n\theta)^{n-1}} \quad (\text{A.21})$$

for $x = 0, 1, 2, \dots, n$ and zero otherwise. where $a > 0$, $b > 0$ and $\theta > -a/n$. However, Alzaid and Al-Osh (1993), when defining the QB thinning operator, used a particular case of the QB-II distribution:

$$P(X = x) = pq \binom{n}{x} \frac{(p + x\theta)^{x-1} (q + (n-x)\theta)^{n-x-1}}{(1 + n\theta)^{n-1}} \quad (\text{A.22})$$

where $a = p$, $q = b$, $0 < q = 1 - p < 1$ and θ is such that $n\theta < \min(p, q)$. We denote the QB-II with $\text{QB}(p, \theta, n)$. For large n , such that $np \rightarrow \lambda$, the QB distribution tends to the Generalized Poisson distribution.

The quasi-binomial (QB) thinning has been introduced by Alzaid and Al-Osh (1993) as a generalization of the binomial thinning to model processes with GP distribution marginals. Unlike the binomial thinning, the QB thinning is able to obtain the distribution of the corresponding innovation. In particular, the authors argued in many counting process is more suitable to consider that the probability of retaining an element may depend on the time and/or the number of already retained elements. They assumed that, at time t , the number of retained elements follows a QB distribution. Using the notation in Weiß (2008), the QB thinning is defined as follows:

Proposition 11 (Quasi-Binomial Thinning). *Let $\rho_{\theta, \lambda} \circ$ be the quasi-binomial thinning operator such that $\rho_{\theta, \lambda} \circ$ follows a $\text{QB}(\rho, \theta / \lambda, x)$. If X follows a $\text{GP}(\lambda, \theta)$ distribution and the quasi-binomial thinning is performed independently on X , then $\rho_{\theta, \lambda} \circ X$ has a $\text{GP}(\rho\lambda, \theta)$ distribution.*

The thinned variable, $\rho_{\theta, \lambda} \circ X$, can be interpreted as the number of survivors from a population described by X .

Properties

- Expected value: $E[\rho_{\theta, \lambda} \circ X] = \rho \cdot \mu_X$
- Covariance: $E[\rho_{\theta, \lambda} \circ X, X] = \rho \cdot \sigma_X^2$

B Proofs of the results of the paper

B.1 Convolution property of the Generalized Poisson distribution in Section A

Proof of Lemma 1. Let $X \sim \text{GPD}(\theta_1, \theta_2, \lambda)$ and $Y \sim \text{GPD}(\theta_3, \theta_4, \lambda)$. We can write each r.v. as

$$X = X_1 - X_2 \text{ and } Y = Y_1 - Y_2 \quad (\text{B.1})$$

where $X_1 \sim GP(\theta_1, \lambda)$, $X_2 \sim GP(\theta_2, \lambda)$, $Y_1 \sim GP(\theta_3, \lambda)$ and $Y_2 \sim GP(\theta_4, \lambda)$. Therefore we can write

$$\begin{aligned} X + Y &= X_1 - X_2 + Y_1 - Y_2 \\ &= (X_1 + Y_1) - (X_2 + Y_2) \\ &= GP(\theta_1 + \theta_3, \lambda) - GP(\theta_2 + \theta_4, \lambda) \sim GPD(\theta_1 + \theta_3, \theta_2 + \theta_4, \lambda). \end{aligned} \quad (\text{B.2})$$

We can generalize the result as follows. Let $Z_i \stackrel{\text{i.i.d.}}{\sim} GPD(\theta_1, \theta_2, \lambda)$. Then we have

$$\sum_{i=1}^n Z_i \sim GPD\left(\sum_{i=1}^n \theta_{1i}, \sum_{i=1}^n \theta_{2i}, \lambda\right). \quad (\text{B.3})$$

In the same way we can prove that the difference of two r.v. GPD distributed, is again a GPD.

Let $X \sim GPD(\theta_1, \theta_2, \lambda)$ and $Y \sim GPD(\theta_3, \theta_4, \lambda)$. We can write each r.v. as

$$X = X_1 - X_2 \text{ and } Y = Y_1 - Y_2 \quad (\text{B.4})$$

where $X_1 \sim GP(\theta_1, \lambda)$, $X_2 \sim GP(\theta_2, \lambda)$, $Y_1 \sim GP(\theta_3, \lambda)$ and $Y_2 \sim GP(\theta_4, \lambda)$. Therefore we can write

$$\begin{aligned} X - Y &= X_1 - X_2 - Y_1 + Y_2 \\ &= (X_1 + Y_2) - (X_2 + Y_1) \\ &= GP(\theta_1 + \theta_4, \lambda) - GP(\theta_2 + \theta_3, \lambda) \sim GPD(\theta_1 + \theta_4, \theta_2 + \theta_3, \lambda). \end{aligned} \quad (\text{B.5})$$

□

B.2 Proofs of the results in Section 2

Proof of Remark 1. Let $X \sim GP(\theta_1, \lambda)$ and $Y \sim GP(\theta_2, \lambda)$, then $Z = (X - Y)$ follows a $GPD(\theta_1, \theta_2, \lambda)$. We know that $P(Z = z) = P(X = x) \cdot P(Y = y)$, therefore we can name $S = Y$ and we substitute S in $Z = X - Y$, obtaining $X = S + Z$. Now we can write:

$$P(Z = z) = e^{-(\theta_1 + \theta_2)} \cdot e^{-z\lambda} \sum_{s=\max(0, -z)}^{+\infty} \frac{\theta_1 \theta_2}{s!(s+z)!} (\theta_2 + \lambda s)^{s-1} (\theta_1 + \lambda(s+z))^{s+z-1} \cdot e^{-2\lambda s} \quad (\text{B.6})$$

which is the probability of a $GPD(\theta_1, \theta_2, \lambda)$. We can now introduce the new parametrization of the probability density function of the GPD. Define

$$\begin{cases} \mu = \theta_1 - \theta_2 \\ \sigma^2 = \theta_1 + \theta_2 \end{cases}, \quad (\text{B.7})$$

thus, we can rewrite both parameters θ_i , $i = 1, 2$, with respect to μ and σ^2 :

$$\begin{cases} \theta_1 = \frac{\sigma^2 + \mu}{2} \\ \theta_2 = \frac{\sigma^2 - \mu}{2} \end{cases} \quad (\text{B.8})$$

By substituting B.8 into equation B.6 we have

$$\begin{aligned} P(Z = z) &= e^{(\frac{\sigma^2 + \mu}{2} + \frac{\sigma^2 - \mu}{2})} e^{-z\lambda} \sum_{s=\max(0, -z)}^{+\infty} \frac{\left(\frac{\sigma^2 + \mu}{2}\right) \left(\frac{\sigma^2 - \mu}{2}\right)}{s!(s+z)!} \left[\frac{\sigma^2 + \mu}{2} + (s+z)\lambda\right]^{s+z-1} \\ &\quad \left[\frac{\sigma^2 - \mu}{2} + s\lambda\right]^{s-1} e^{-2\lambda s} \end{aligned} \quad (\text{B.9})$$

Carrying out the operations in Eq. B.9 we obtain Eq. B.7. \square

Proof of Remark 2. If $Z \sim GPD(\theta_1, \theta_2, \lambda)$ in the parametrization of Consul (1986), the moments are given in equations A.17-A.20. By using our reparametrization of the GDP

$$\theta_1 = \frac{\sigma^2 + \mu}{2}, \quad \theta_2 = \frac{\sigma^2 - \mu}{2} \quad (\text{B.10})$$

in equations A.17-A.20, we obtain

$$E(Z) = \frac{\theta_1 - \theta_2}{1 - \lambda} = \frac{\mu}{1 - \lambda} \quad (\text{B.11})$$

$$V(Z) = \frac{\theta_1 + \theta_2}{(1 - \lambda)^3} = \frac{\sigma^2}{(1 - \lambda)^3} = \kappa_2 \quad (\text{B.12})$$

$$\begin{aligned} S(Z) &= \frac{\mu^{(3)}}{\sigma^3} = \frac{\kappa_3}{\kappa_2^{3/2}} \\ &= \frac{\mu(1 + 2\lambda)}{(1 - \lambda)^5} \left(\frac{(1 - \lambda)^3}{\sigma^2}\right)^{3/2} \\ &= \frac{\mu(1 + 2\lambda)}{\sigma^3 \sqrt{1 - \lambda}} \end{aligned} \quad (\text{B.13})$$

$$\begin{aligned} K(Z) &= \frac{\kappa_4 + 3\kappa_2^2}{\sigma^4} \\ &= \left(\frac{\sigma^2(1 + 8\lambda + 6\lambda^2)}{(1 - \lambda)^7} + \frac{3\sigma^4}{(1 - \lambda)^6}\right) \left(\frac{(1 - \lambda)^6}{\sigma^4}\right) \\ &= 3 + \frac{1 + 8\lambda + 6\lambda^2}{\sigma^2(1 - \lambda)} \end{aligned} \quad (\text{B.14})$$

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

$$\begin{aligned}
\mu &= E[Z_t] \\
&= E[E[Z_t|\mathcal{F}_{t-1}]] \\
&= E\left[\alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j Z_{t-j}\right] \\
&= \alpha_0 D^{-1}(1) + \mu D^{-1}(1)G(1)
\end{aligned} \tag{B.15}$$

$$\Rightarrow \mu = \frac{\alpha_0}{D(1) - G(1)} = \alpha_0 \left(1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j\right)^{-1} = \alpha_0 K^{-1}(1). \tag{B.16}$$

Where we go from line two to line three of eq B.15 as follows

$$\begin{aligned}
E[Z_t|\mathcal{F}_{t-1}] &= E[\alpha_0 D^{-1}(1) + H(B)Z_t|\mathcal{F}_{t-1}] \\
&= E[\alpha_0 D^{-1}(1)] + E[H(B)Z_t|\mathcal{F}_{t-1}] \\
&= \alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j Z_{t-j}.
\end{aligned} \tag{B.17}$$

Following Ferland et al. (2006), to go from line three to line four of B.15:

$$\begin{aligned}
E[\alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j Z_{t-j}] &= E[\alpha_0 D^{-1}(1)] + E[\sum_{j=1}^{\infty} \psi_j Z_{t-j}] \\
&= \alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j E[Z_{t-j}] \\
&= \alpha_0 D^{-1}(1) + \mu H(1) \\
&= \alpha_0 D^{-1}(1) + \mu D^{-1}(1)G(1).
\end{aligned} \tag{B.18}$$

From B.16, a necessary condition for the second-order stationarity of the integer-valued process $\{Z_t\}$ is: $\left(1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j\right) > 0$. \square

B.3 Proof of the results in Section 3

Proof of Proposition 3.

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

□

Proof of Proposition 4. In order to prove the almost sure convergence of $\{Z_t^{(n)}\}$ we will prove that the difference of two sequences $\{X_t^{(n)}\}$ and $\{Y_t^{(n)}\}$ that have an almost sure convergence, will have an almost sure convergence.

We know that $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$, where $X_t^{(n)}$ and $Y_t^{(n)}$ are two sequences of GP random variable. From Zhu (2012) we have

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

and

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

Let

$$A = \{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}$$

and

$$B = \{\omega \in \Omega \times \Omega : \lim_{n \rightarrow \infty} (aX_n(\omega) + bY_n(\omega)) = aX(\omega) + bY(\omega)\}, \quad \forall a, b \in \mathbb{R}.$$

Now we show the almost sure convergence of the sum $(aX_n(\omega) + bY_n(\omega))$.

$$\begin{aligned}
\int_{\Omega} \mathbb{I}_B(\omega) d\mathbb{P}(\omega) &= \int_{(\Omega \cap A) \cup (\Omega \cap A^c)} \mathbb{I}_B(\omega) d\mathbb{P}(\omega) \\
&= \int_{\Omega} \mathbb{I}_B(\omega) \mathbb{I}_A(\omega) d\mathbb{P}(\omega) + \int_{\Omega} \mathbb{I}_B(\omega) \mathbb{I}_{A^c}(\omega) d\mathbb{P}(\omega) \\
&= \int_{\Omega} \mathbb{I}_B(\omega) \mathbb{I}_A(\omega) d\mathbb{P}(\omega) + \int_{\Omega} \mathbb{I}_{A^c}(\omega) \left(\int_{\Omega} \mathbb{I}_{B \cap A^c}(\omega) d\mathbb{P}(\omega | \omega') \right) d\mathbb{P}(\omega') \\
&= \mathbb{P}(B|A) \mathbb{P}(A) + \underbrace{\mathbb{P}(B|A^c) \mathbb{P}(A^c)}_{=0} \\
&= \mathbb{P}(\{\omega : \lim_{n \rightarrow \infty} (aX_n(\omega) + bY_n(\omega)) = aX(\omega) + bY(\omega)\} | A) \underbrace{\mathbb{P}(A)}_{=1} \\
&= \mathbb{P}(\{\omega : a \lim_{n \rightarrow \infty} X_n(\omega) = aX(\omega) - bY(\omega) + bY(\omega)\}) = 1
\end{aligned} \tag{B.19}$$

Therefore, if $X_n^{(\omega)} \xrightarrow{\text{a.s.}} X(\omega)$ and $Y_n^{(\omega)} \xrightarrow{\text{a.s.}} Y(\omega)$

$$\Rightarrow aX_n(\omega) + bY_n(\omega) \xrightarrow{\text{a.s.}} aX(\omega) + bY(\omega), \quad \forall a, b \in \mathbb{R}.$$

$\forall a, b \in \mathbb{R}$. Hence, for $a = 1$ and $b = -1$, this is true for the difference $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$. \square

Proof of Proposition 5. We use again the fact that $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$ and the following lemma.

Lemma 2. *If $X_n(\omega)$ and $Y_n(\omega)$ have a mean-square limit*

$$\begin{aligned}
X_n(\omega) &\xrightarrow{L^2} X(\omega) \\
Y_n(\omega) &\xrightarrow{L^2} Y(\omega)
\end{aligned} \tag{B.20}$$

also their sum will have a mean-square limit.

$$aX_n(\omega) + bY_n(\omega) \xrightarrow{L^2} aX(\omega) + bY(\omega), \quad \forall a, b \in \mathbb{R} \tag{B.21}$$

Hence, by setting $a = 1$ and $b = -1$ we will obtain that Lemma 2 will be valid also for the difference of two sequences

$$X_n(\omega) - Y_n(\omega) \xrightarrow{L^2} X(\omega) - Y(\omega) \tag{B.22}$$

and we can say that $Z_t^{(n)}$ converges to Z_t in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. \square

Proof of Proposition 6.

$$g_{\mathbf{Z}}(\mathbf{t}) = \mathbb{E} \left[\prod_{i=1}^k t_i^{Z_i} \right] = \mathbb{E} \left[\prod_{i=1}^k t_i^{X_i} \right] \mathbb{E} \left[\prod_{i=1}^k \frac{1}{t_i^{Y_i}} \right] = g_{\mathbf{X}}(\mathbf{t})g_{\mathbf{Y}}(\mathbf{t}^{-1}) \quad (\text{B.23})$$

□

Proof of Proposition 9. As said before, $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$. Where $X_t^{(n)}$ and $Y_t^{(n)}$ are finite sums of independent Generalized Poisson variables and it follows that $Z_t^{(n)}$ 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: $E[X_t] = \mu_X \leq C_1$, $E[Y_t] = \mu_Y \leq C'_1$, $V[X_t] = \sigma_X^2 \leq C_2$, $V[Y_t] = \sigma_Y^2 \leq C'_2$, therefore,

$$E[Z_t] = E[X_t] - E[Y_t] = \mu_X - \mu_Y \leq \mu_X + \mu_Y \leq C_1 + C'_1 \quad (\text{B.24})$$

is finite and

$$V[X_t - Y_t] = V[X_t] + V[Y_t] = \sigma_X^2 + \sigma_Y^2 \leq C_2 + C'_2 \quad (\text{B.25})$$

is also finite, where $Cov(X_t, Y_t) = 0$ given that X_t and Y_t are independent and where C_i and C'_i , with $i = 1, 2$ are constants. □

Proof of the results in Example 1. For $k \geq 2$:

$$\gamma_Z(k) = \alpha_1 \gamma_Z(k-1) = \alpha_1^{k-1} \gamma_Z(1) \quad (\text{B.26})$$

For $k = 1$:

$$\gamma_Z(1) = Cov(Z_t, Z_{t-1}) = \alpha_1 \gamma_Z(0) = \alpha_1 V(Z_t) = \alpha_1 [\phi^3 E(\sigma_t^{2*})] + \alpha_1 V(\mu_t). \quad (\text{B.27})$$

For $k \geq 1$ we have

$$\gamma_\mu(k) = \alpha_1 \gamma_\mu(k-1) = \alpha_1^k V(\mu_t). \quad (\text{B.28})$$

For $k = 0$:

$$\begin{aligned} \gamma_\mu(k) &= V(\mu_t) = \alpha_1 \gamma_Z(1) \\ &= \alpha_1 \{ \alpha_1 [\phi^3 E(\sigma_t^{2*})] + V(\mu_t) \} = \alpha_1^2 [\phi^3 E(\sigma_t^{2*})] + \alpha_1^2 V(\mu_t) \end{aligned} \quad (\text{B.29})$$

Therefore,

$$V(\mu_t) = \frac{\alpha_1^2 [\phi^3 E(\sigma_t^{2*})]}{1 - \alpha_1^2} \quad (\text{B.30})$$

and

$$\begin{aligned} V(Z_t) &= \phi^3 E(\sigma_t^{2*}) + V(\mu_t) \\ &= \phi^3 E(\sigma_t^{2*}) + \frac{\alpha_1^2 [\phi^3 E(\sigma_t^{2*})]}{1 - \alpha_1^2} = \frac{\phi^3 E(\sigma_t^{2*})}{1 - \alpha_1^2} \end{aligned} \quad (\text{B.31})$$

where $\phi = \frac{1}{1-\lambda}$. Finally, the autocorrelations are derived as follows:

$$\rho_\mu(k) = \frac{\gamma_\mu(k)}{V(\mu_t)} = \frac{\alpha_1^k V(\mu_t)}{V(\mu_t)} = \alpha_1^k \quad (\text{B.32})$$

$$\begin{aligned} \rho_Z(k) &= \frac{\gamma_Z(k)}{V(Z_t)} = \alpha_1^{k-1} \gamma_Z(1) \frac{1 - \alpha_1^2}{\phi^3 E(\sigma_t^{2*})} \\ &= \alpha_1^{k-1} \frac{\alpha_1(1 - \alpha_1^2)\phi^3 E(\sigma_t^{2*}) + \alpha_1^3 \phi^3 E(\sigma_t^{2*})}{1 - \alpha_1^2} \frac{1 - \alpha_1^2}{\phi^3 E(\sigma_t^{2*})} = \alpha_1^k \end{aligned} \quad (\text{B.33})$$

□

Proof of the results in Example 2. For $k \geq 2$:

$$\gamma_Z(k) = \alpha_1 \gamma_Z(k-1) + \beta_1 \gamma_Z(k-1) = (\alpha_1 + \beta_1)^{k-1} \gamma_Z(1) \quad (\text{B.34})$$

For $k = 1$:

$$\begin{aligned} \gamma_Z(1) &= \text{Cov}(Z_t, Z_{t-1}) = \alpha_1 \gamma_Z(0) + \beta_1 \gamma_\mu(0) \\ &= \alpha_1 V(Z_t) + \beta_1 V(\mu_t) = \alpha_1 [\phi^3 E(\sigma_t^{2*})] + (\alpha_1 + \beta_1) V(\mu_t). \end{aligned} \quad (\text{B.35})$$

For $k \geq 1$ we have

$$\gamma_\mu(k) = \alpha_1 \gamma_\mu(k-1) + \beta_1 \gamma_\mu(k-1) = (\alpha_1 + \beta_1)^k V(\mu_t). \quad (\text{B.36})$$

For $k = 0$:

$$\begin{aligned} \gamma_\mu(k) &= V(\mu_t) = \alpha_1 \gamma_Z(1) + \beta_1 \gamma_\mu(1) \\ &= \alpha_1 \{ \alpha_1 [\phi^3 E(\sigma_t^{2*})] + (\alpha_1 + \beta_1) V(\mu_t) \} + \beta_1 [(\alpha_1 + \beta_1) V(\mu_t)] \\ &= \alpha_1^2 [\phi^3 E(\sigma_t^{2*})] + (\alpha_1 + \beta_1)^2 V(\mu_t) \end{aligned} \quad (\text{B.37})$$

Therefore,

$$V(\mu_t) = \frac{\alpha_1^2 [\phi^3 E(\sigma_t^{2*})]}{1 - (\alpha_1 + \beta_1)^2} \quad (\text{B.38})$$

and

$$\begin{aligned} V(Z_t) &= \phi^3 E(\sigma_t^{2*}) + V(\mu_t) = \phi^3 E(\sigma_t^{2*}) + \frac{\alpha_1^2 [\phi^3 E(\sigma_t^{2*})]}{1 - (\alpha_1 + \beta_1)^2} \\ &= \frac{\phi^3 E(\sigma_t^{2*}) [1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2]}{1 - (\alpha_1 + \beta_1)^2} \end{aligned} \quad (\text{B.39})$$

where $\phi = \frac{1}{1-\lambda}$. The autocorrelations are derived as follows:

$$\rho_\mu(k) = \frac{\gamma_\mu(k)}{V(\mu_t)} = \frac{(\alpha_1 + \beta_1)^k V(\mu_t)}{V(\mu_t)} = (\alpha_1 + \beta_1)^k \quad (\text{B.40})$$

$$\begin{aligned}
\rho_Z(k) &= \frac{\gamma_Z(k)}{V(Z_t)} = (\alpha_1 + \beta_1)^{k-1} \gamma_Z(1) \frac{1 - (\alpha_1 + \beta_1)^2}{\phi^3 E(\sigma_t^{2*}) [1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2]} \\
&= (\alpha_1 + \beta_1)^{k-1} \frac{\alpha_1 [1 - \beta_1(\alpha_1 + \beta_1)]}{1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2}
\end{aligned} \tag{B.41}$$

□

C Numerical Illustration

We consider 400 samples from a two GPD-INGARCH(1,1) and two 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 finally 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 Fig. C.1-C.3 the MCMC output for one MCMC draw before removing the burn-in sample and thinning, while in Fig.C.4-C.6 the MCMC output after removing the burn-in sample and thinning.

C.1 Before thinning

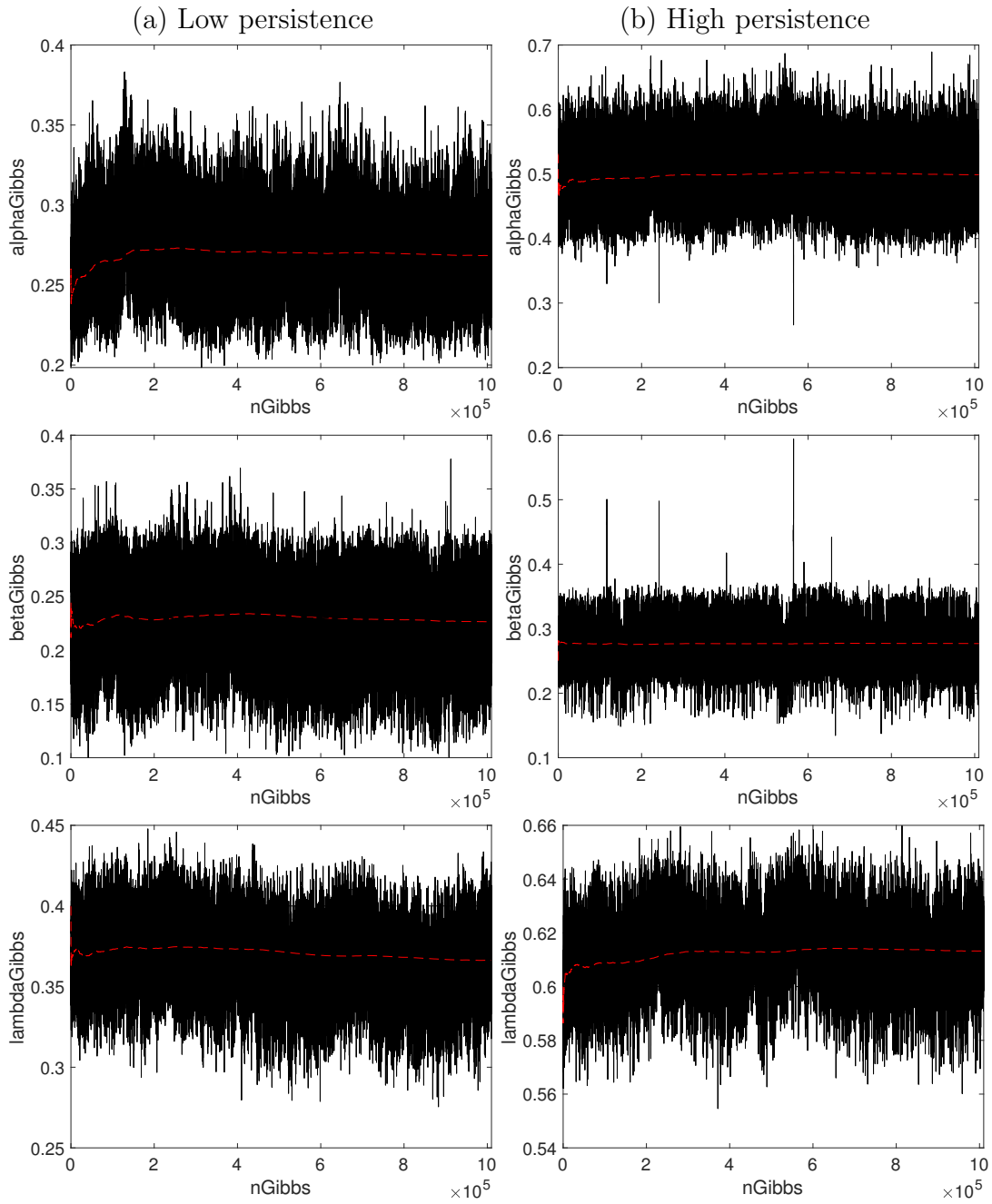


Figure C.1: MCMC plot for the parameters in the two setting: low persistence and high persistence.

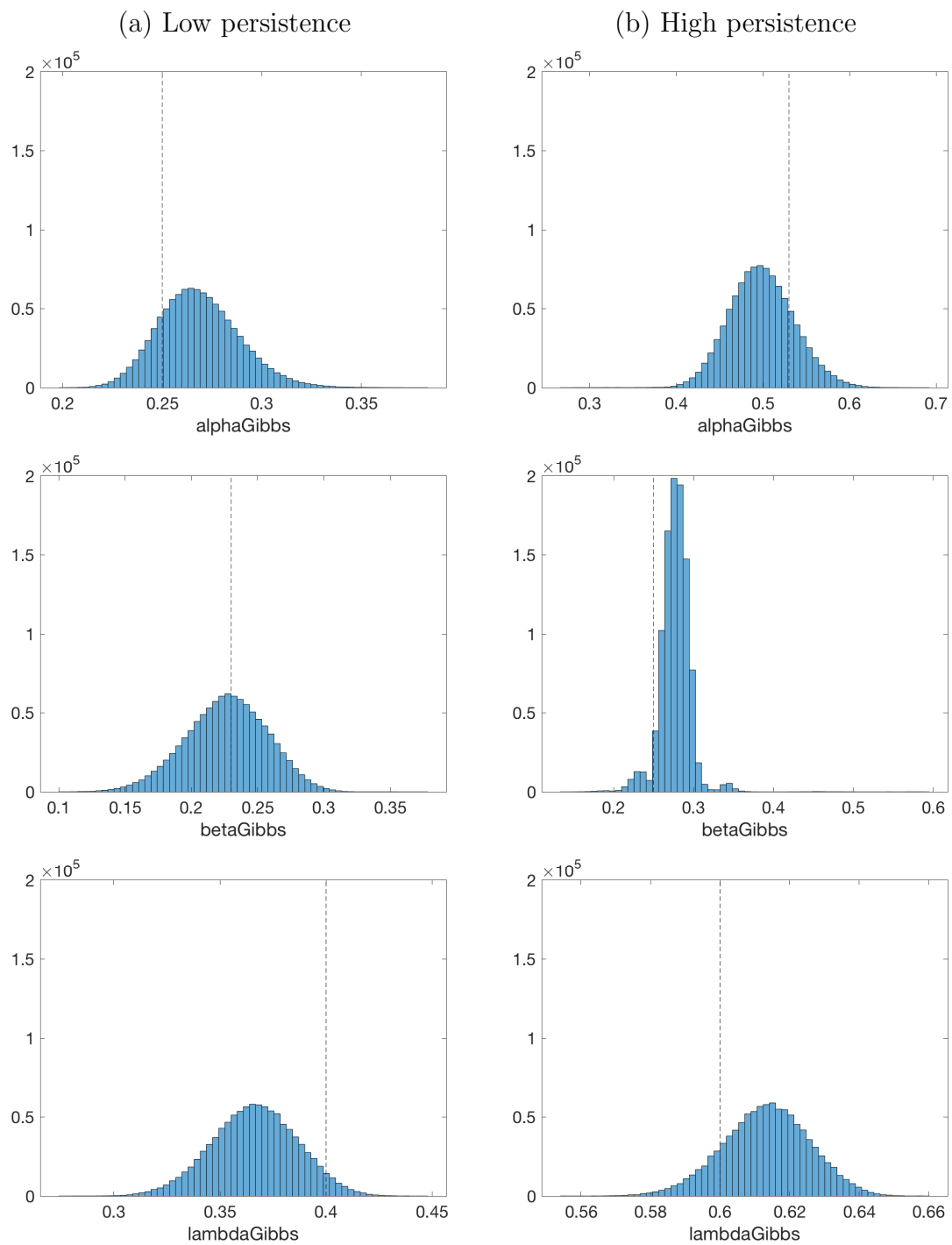


Figure C.2: Histograms of the MCMC draws for the parameters in both settings: low persistence and high persistence.

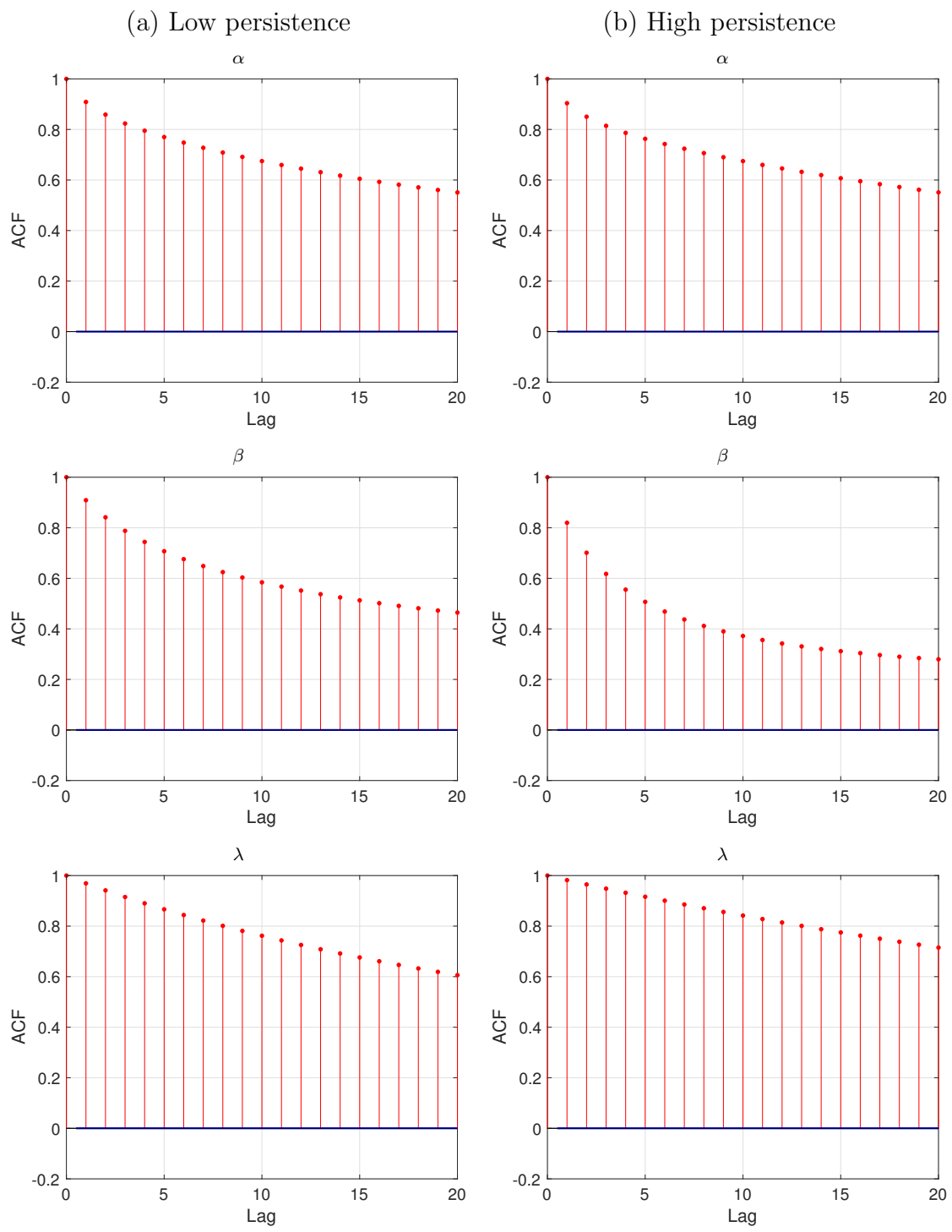


Figure C.3: Autocorrelation function for the parameters in both low persistence and high persistence settings.

C.2 After thinning

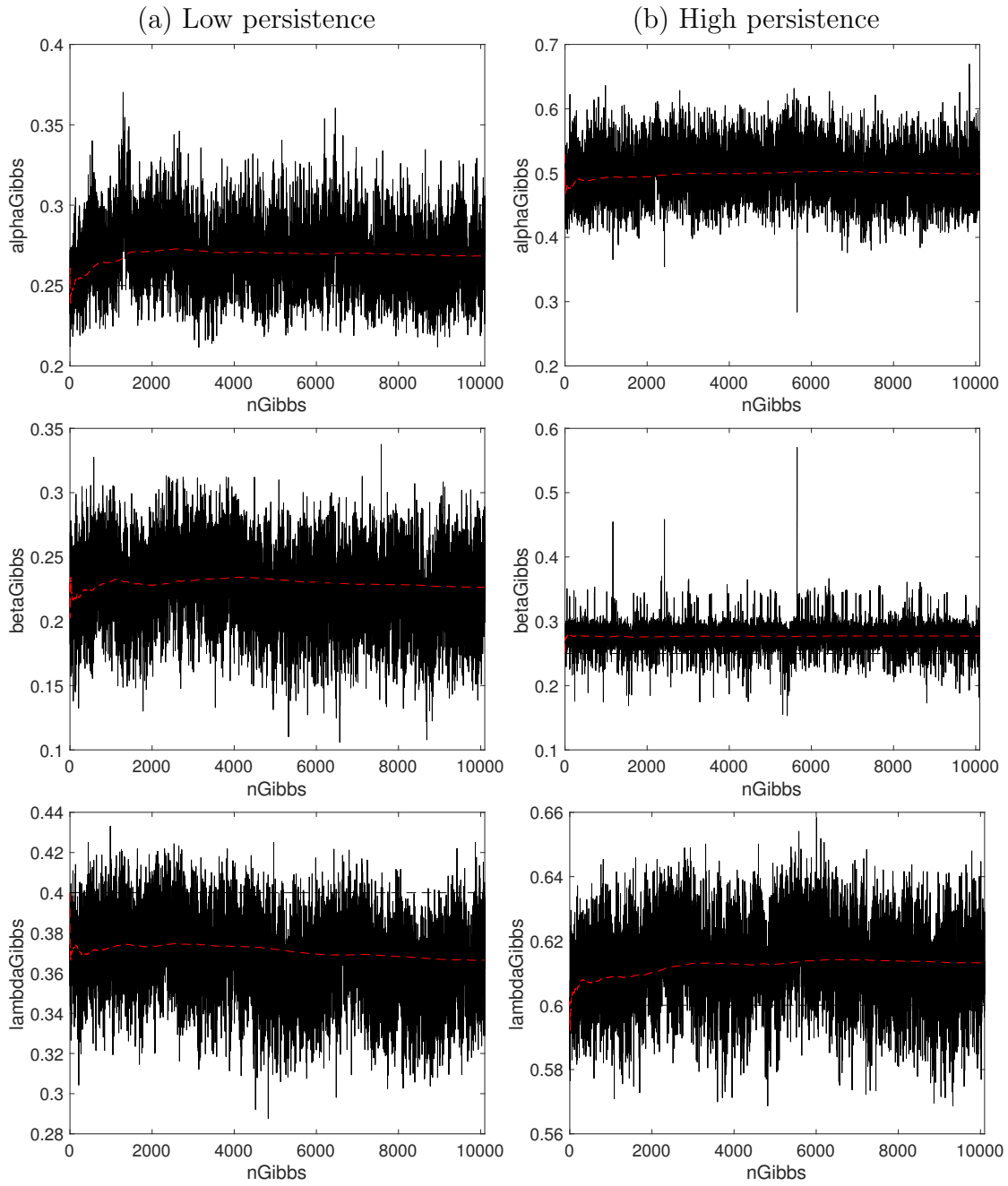


Figure C.4: MCMC plot for the parameters in the two setting: low persistence and high persistence, after removing the burn-in sample and thinning.

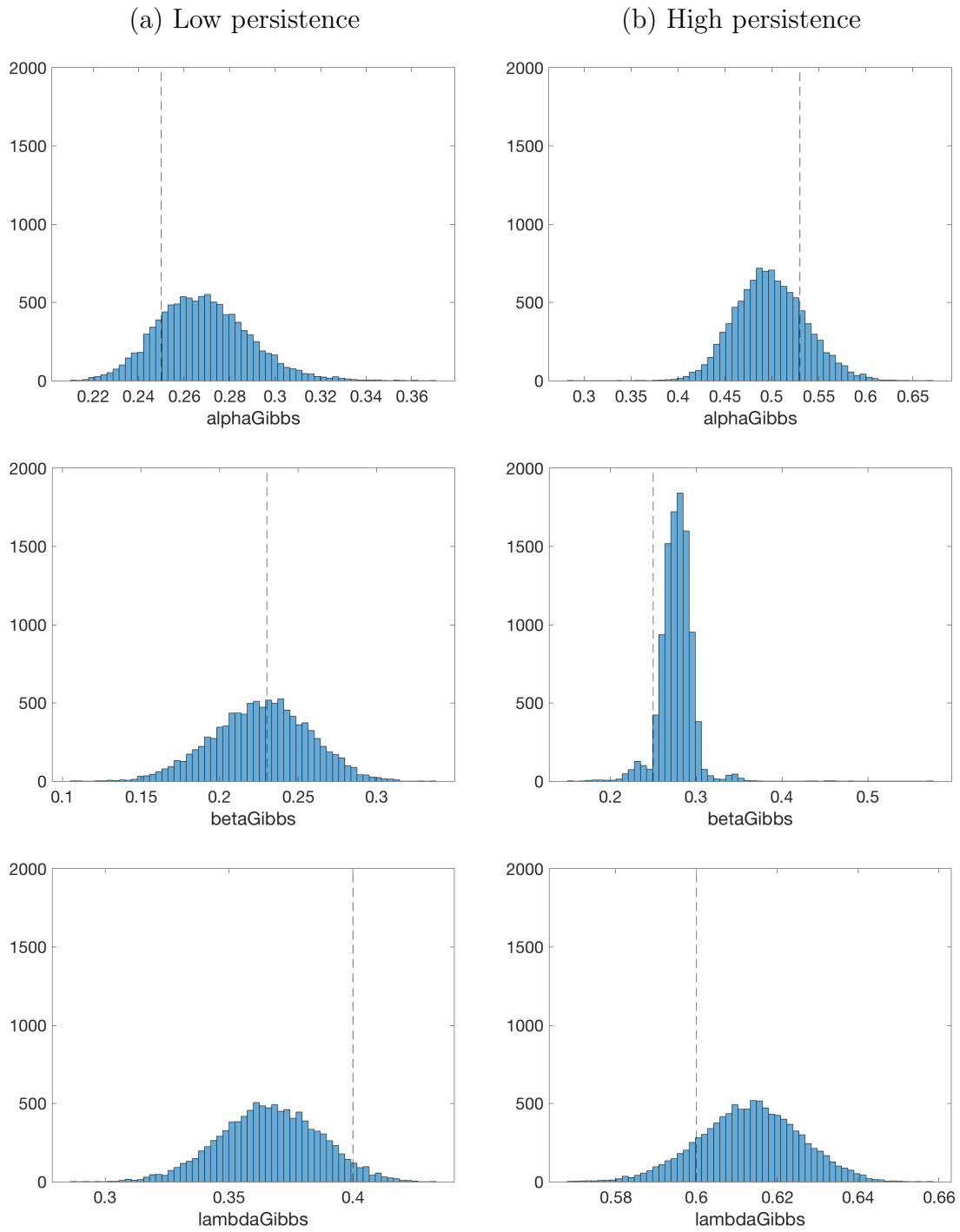


Figure C.5: Histograms of the MCMC draws for the parameters in both settings: low persistence and high persistence, after removing the burn-in sample and thinning.

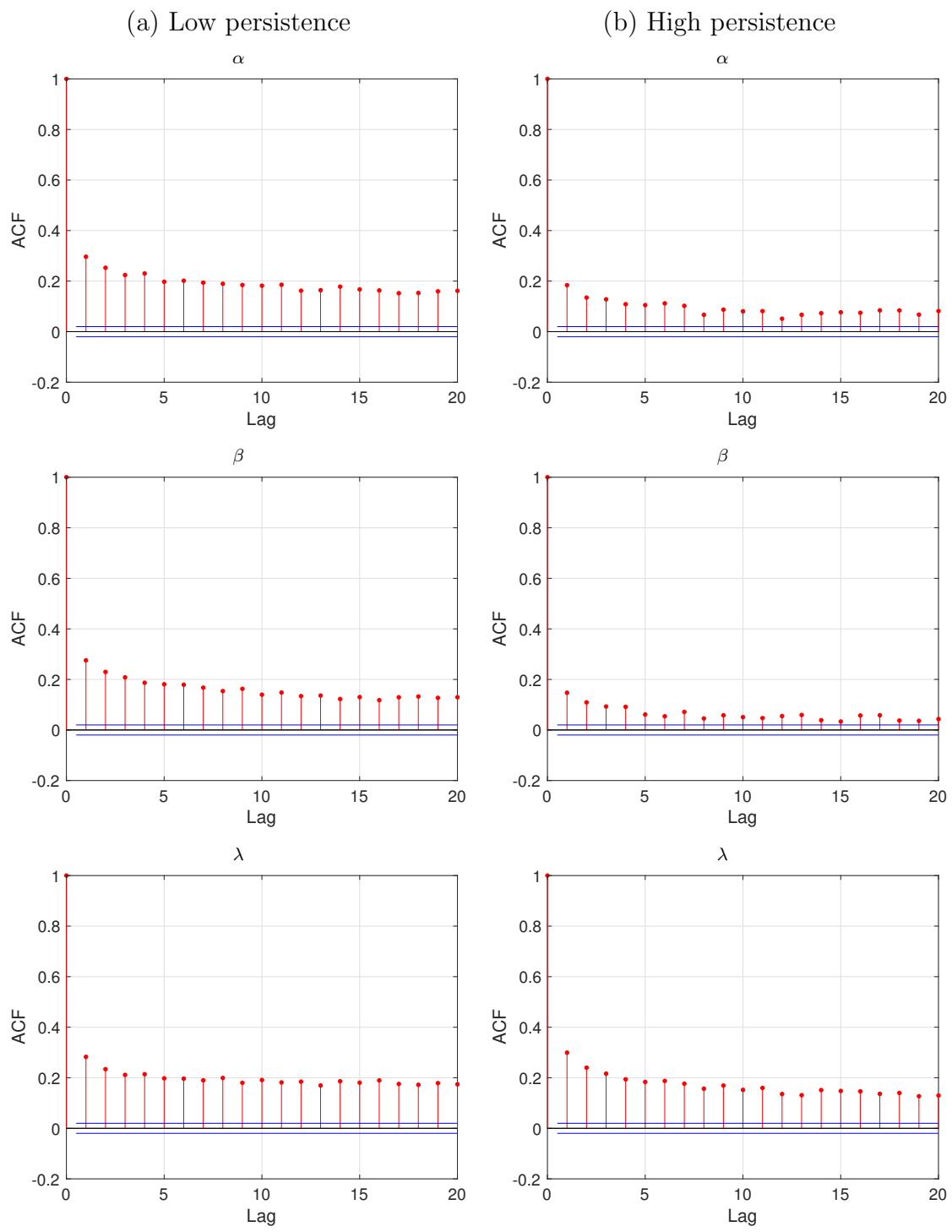


Figure C.6: Autocorrelation function for the parameters in both low persistence and high persistence settings, after removing the burn-in sample and thinning.