Università
Ca'Foscari
Venezia

Corso di Dottorato di Ricerca

in Economia

Ciclo XXXIII

Tesi di Ricerca

Modelling Times Series of Counts

SSD: SECS-P/05

Coordinatore del Dottorato
prof. Pietro Dindo
Supervisore
ch. prof. Roberto Casarin
Secondo supervisore
ch. prof. Monica Billio

Dottoranda
Giulia Carallo
Matricola 956387

The undersigned Giulia Carallo, in her quality of doctoral candidate for a Ph.D. degree in Economics granted by the Università Cà Foscari di Venezia attests that the research exposed in this dissertation is original and that it has not been and it will not be used to pursue or attain any other academic degree of any level at any other academic institution, be it foreign or Italian.

To my soulmate and my grandmother

Acknowledgements

As always, writing this acknowledgments will be difficult for me. I was never good in thanking people, but I will put as much effort as I can.

Firstly, I want to thank my first supervisor Roberto Casarin. When I started this journey, in 2017, I was focused in other econometric topics but Roberto was able to make me interested in the field of this thesis. Working with him pushed me to do always better.

I want to thank also my second supervisor Monica Billio, her accomplishments have been an example to follow and to pursue.

I also thank Christian Robert, Dario Palumbo e Federico Bassetti. I worked with each of them during the realisation of this thesis, and they all taught me a lot about doing research. How to solve the ups and downs you are confronted when starting or at the end of a research.

I also want to thank all my colleagues and the people I met during the PhD. We had very (very!!!!) hard times, we had happy moments and shared one hell of a journey.

Thanks to my parents. You have been next to my during all my studying career, supporting all my choices, helping every time I needed and most importantly always believing in what I was doing. You never forced me to do something. You have been my rationality when I was too tired or when I couldnt see things in the right perspectives.

Thanks to my friends. I know I wasnt the best of the friends during this period, especially the last two years. But youre always there, and I am so grateful of having you in my life.

When I started this PhD the last, emphasis on last, thing that I wanted or thought was to found love. I never thought that a coffee with an old friend wouldve led me to my soulmate to whom this thesis is dedicated. Thank you Alberto, this journey wouldve been very different without you.

This thesis is also dedicated to someone that is not with me anymore and that I wish to thank, my grandmother. She never understood what I was doing in practice, but she was always the first to ask how I was doing, to call me at every times and I regret not answering a lot of times. Thank you gran.

Unfortunately, in the middle of my journey the world has faced one of the biggest pandemic in the recent years. I cannot write my thanks without mentioning this episode, since it affected my academic and personal life. For me, it has been difficult continuing my normal life during this period. Working has been hard, but in the end I was able to overcome the psychological downfall of this pandemic, and now I am here at the end of this journey. Hence, maybe I will seem selfish, but my last thanks go to myself. You made it.

Abstract

This thesis introduces new stochastic process with values in the set of integers. We introduce three different process to model the dynamics in the over-dispersion and under-dispersion feature in the data. In Chapter 1 we introduce an application of Generalized Poisson models for analysing over-dispersion in cyber-attacks. This chapter is motivational for the Generalised Poisson difference INGARCH model introduced in Chapter 2. In Chapter 3, we provide a generalisation defining a new Dynamic Conditional Score process with data distributed as Generalised Poisson. Finally, in Chapter 4 we introduce a new generalisation regarding the innovations. The model introduced is a Generalised Lagrangian Katz INAR process.

Contents

List of Figures

List of Tables

Introduction

The increased availability of count data has led to greater interest in the study of integer-valued models. Since the first econometric model of Hausman et al. (1984), namely the Poisson Regression Model (PRM), different models have been introduced in various directions. In particular, extensions have been studied in order to overcome the Poisson distribution's equidispersion. Cameron and Trivedi (1986) and Lee (1986) contributed by introducing new tests for overdispersion in the data. Since then, it has been noticed that many count data series are not equally dispersed, and Binomial or Negative binomials marginal distributions have been used in order to introduce over and under-dispersion. This thesis introduces new processes to model times series of counts with over or under-dispersion. All chapters are original papers.

Chapter 1 presents an application of Generalized Poisson models for analyzing over-dispersion in cyber-attacks following a Bayesian inference approach, and we apply a Markov Chain Monte Carlo algorithm for posterior approximation. This case study shows the importance of introducing new dynamic models able to capture the dispersion feature of count data.

Chapter 2 introduces a new dynamic model for integer-valued data with sign. Namely, the model is an Integer-Valued GARCH, and the increments are Generalised Poisson difference distributed. For this process, we studied its properties and provided a Bayesian inference framework and an efficient posterior approximation procedure based on Markov Chain Monte Carlo, and we show its correctness and effectiveness on simulated data. We proposed two applications on an accident dataset and cyber threats data.

In Chapter 3, we introduce a new dynamic process where we assume a Dynamic Conditional Score (DCS) dynamics. We derived the score and provided a Bayesian inference framework with an efficient Monte Carlo Markov Chain sampler for posterior approximation. We proposed an application on fires data and showed that the model proposed is suited for capturing persistence and dispersion in the data.

Chapter 4 introduces a novel integer-valued autoregressive process proposed with Generalized Lagrangian Katz innovations. Theoretical properties of the

model, such as stationarity, moments, and semi-self-decomposability, are provided. A Bayesian approach to inference is proposed, and an efficient Markov Chain Monte Carlo sampling procedure has been proposed. The simulation experiments show the effectiveness of the MCMC. The application on the public concern about climate change using Google Trend data, shows that the model captures the dispersion feature of the data.

This thesis contributes to the literature on times series for integers and gives new methods for the econometric analysis of such data. It also contributes to the study of climate change, offering new models that can be easily used in practice for estimating and predicting its determinants

Chapter 1

A Bayesian Generalized Poisson Model for Cyber Risk Analysis [1]

Abstract Cyber threats are now considered as a top risk for many economic sectors which include retail, financial services, security, and healthcare. The costs for damages from cyber-attacks and the number of cyber-attacks are two of the main quantities of interest when measuring cyber-risk. In this chapter, we focus on the frequency of cyber-attacks and analyse some features through the lens of a generalized Poisson model. We follow a Bayesian inference approach and apply a Markov Chain Monte Carlo algorithm for posterior approximation. In the application to a well-known dataset on cyber-threats we find evidence of over-dispersion and of time-variations in the features of the phenomenon.

1.1 Introduction

The digital transformation and the increased use of big data and cloud computing make the economic system more vulnerable to cyber attacks. Among the most exposed economic actors, financial and insurance companies are becoming increasingly susceptible to cyber attacks since they possess substantial amounts of confidential customers information. These facts make cyber-risk one of the most relevant risks in finance and insurance (e.g., see EIOPA (2018),EIOPA (2019)).

On the other hand, the digital economy and the technology advances also offer opportunities to the underwriters of cyber insurance policies EIOPA (2019). A well-developed cyber insurance market can play a crucial role in favouring the transformation to the digital economy, by raising awareness of cyber risks, sharing knowledge on good cyber risk management practices and facilitating responses to and recovery from cyber attacks.

[1] In collaboration with Christian P. Robert (Université Paris-Dauphine, France and University of Warwick, UK). This chapter appeared as Carallo, G., Casarin, R., and Robert, C. (2022), A Bayesian Generalized Poisson Model for Cyber Risk Analysis in Corazza, M., Gilli, M., Perna, C., Pizzi, C., Sibillo, M. (eds.), Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer Verlag.

The number of cyber-attacks is one of the relevant variables to model and predict when measuring and managing the costs for damages originated by cyber events. Despite the relevance of cyber risks, a few works focused on the statistical modelling of cyber-attacks (e.g., see Chen and Lee (2016) Chen et al. (2016), Xu et al. (2017)). Also, see Husák et al. (2018) for a review on statistical, game-theoretic and machine learning methods for predicting cyber-attacks. In this chapter, we focus on Poisson models which are well suited for integer-valued data, such as the number of cyber-attacks. The standard Poisson distribution has been proposed for modelling cyber risk, nevertheless this model is not flexible enough to capture some features of the data which are relevant for a correct evaluation of cyber-risk Leslie et al. (2018). The aim of this chapter is to propose Generalized Poisson models for analyzing over-dispersion in cyber-attacks. We follow a Bayesian inference approach (e.g., see Zhu and Li (2009) and Scollnik (1998)) which allows for including into the inference process extra-sample information, such as opinions of experts. This extra-sample information can be appealing in cyber-risk modelling when a few observations are available.

The chapter is organized as follows: in Section 1.2 we present the model and briefly discuss the inferential approach. Section 1.3 presents the empirical results obtained on a well-known dataset on cyber-threats. Section 1.4 concludes.

1.2 A Bayesian Generalized Poisson Model

The Generalized Poisson (GP) distribution of a random variable X with parameters $\theta > 0$ and $\lambda \in (0, 1)$ is given by

$$\mathbb{P}(X = x|\theta, \lambda) = \frac{\theta(\theta + \lambda x)^{x-1}}{x!} e^{(-\theta-\lambda x)}, \quad x = 0, 1, 2, \ldots \tag{1.1}$$

The GP distribution is part of the general class of Lagrangian distributions (see Consul and Jain (1973)) and has the following moments

$$\mathbb{E}(X) = \frac{\theta}{1-\lambda}, \qquad \mu^{(2)} = \frac{\theta}{(1-\lambda)^3} \tag{1.2}$$

$$\mu^{(3)} = \frac{1+2\lambda}{\sqrt{\theta(1-\lambda)}}, \quad \mu^{(4)} = \frac{1+8\lambda+6\lambda^2}{\theta(1-\lambda)} + 3 \tag{1.3}$$

where $\mu^{(r)} = \mathbb{E}(X - \mathbb{E}(X))^r$. For $\lambda = 0$ one obtains the standard Poisson distribution as special case, whereas for $\lambda \in (0, 1)$ over-dispersion is obtained, that is a greater variability than would be explained by the standard Poisson model.

Following Scollnik (1998) we assume the following independent gamma and beta prior distributions

$$\theta \sim \mathcal{G}a(a, b), \quad \lambda \sim \mathcal{B}e(c, d) \tag{1.4}$$

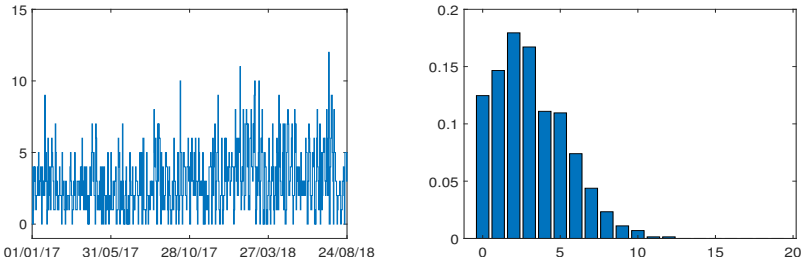


Figure 1.1: Time series (left) and histogram (right) of the total number of cyber threats collected at the daily frequency from 1st January 2018 to 31st December 2018. Data source: `https://www.hackmageddon.com/`.

where a and b are the shape and rate parameters of the gamma distribution. Given a sequence of i.i.d. observations $X_1, \ldots, X_T$ from the GP distribution, the joint posterior distribution of the parameters θ and λ is

$$\pi(\theta, \lambda | X_1, \ldots, X_T) \propto L(X_1, \ldots, X_T | \theta, \lambda) \pi(\theta, \lambda) \tag{1.5}$$

where $L(X_1, \ldots, X_T | \theta, \lambda)$ denotes the likelihood function and $\pi(\theta, \lambda)$ the joint prior distribution given in Eq. 1.4. The posterior distribution is not tractable due to lack of conjugacy, thus we propose a Metropolis-Hastings (MH) algorithm to generate random samples from $\pi(\theta, \lambda | X_1, \ldots, X_T)$ and to approximate the Bayes estimators and all the posterior quantities of interest. At the i-th iteration of the MH sampler a candidate for the parameters is generated from the two independent random walk proposal distributions

$$\theta^* \sim \mathcal{G}a(a^{(i)}, b^{(i)}), \quad \lambda^* \sim \mathcal{B}e(c^{(i)}, d^{(i)}) \tag{1.6}$$

where $b^{(i)} = r/\theta^{(i-1)}$, $a^{(i)} = b^{(i)}\theta^{(i-1)}$, $c^{(i)} = s\lambda^{(i-1)}$, $d^{(i)} = s(1 - \lambda^{(i-1)})$ and $\theta^{(i-1)}$ and $\lambda^{(i-1)}$ denote the previous iteration values of the parameters.

1.3 A Cyber Attacks Dataset

Our dataset contains observations on the number of cyber threats collected at the daily frequency from the 1st January 2018 to the 31st December 2018. The data source is `https://www.hackmageddon.com/`. The observations refer to the following classes of threats: cyber crime, cyber espionage, cyber warfare, and hacktivism. The series of the total number of cyber-threats is given in Figure 1.1.

In the following we consider the time series of the total number of cyber events and the Bayesian model and the Markov Chain Monte Carlo (MCMC) procedure proposed in Section 1.2. We set the gamma prior hyper-parameters $a = 10$, $b = 1$ and the beta prior hyper-parameters $c = 2$ and $d = 2$. We set the scale to $r = 0.01$ and the precision to $s = 100$ in the MH random walk

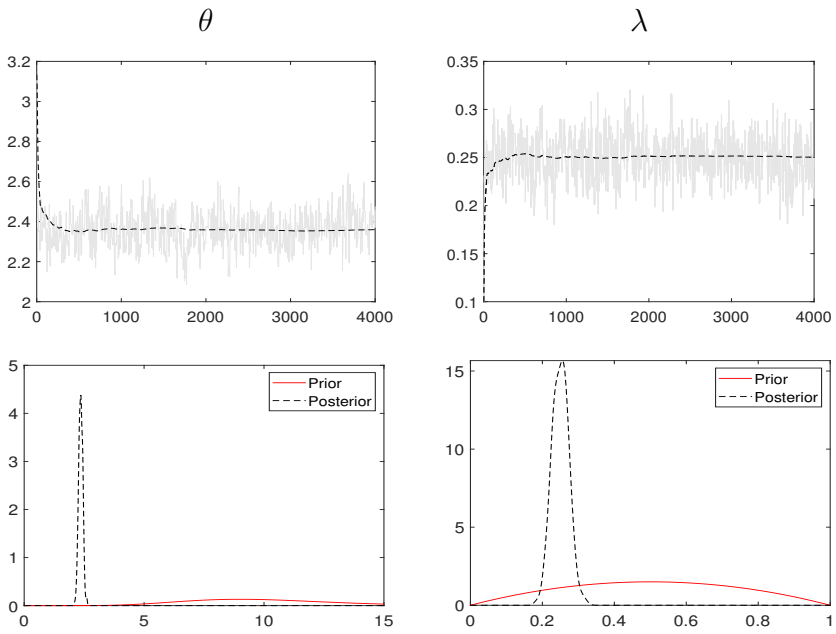


Figure 1.2: Top: posterior MCMC draws (gray solid lines) and progressive averages (black dashed lines) for θ (left) and λ (right). Bottom: prior distributions (red solid lines) and kernel density estimates of the posterior distributions (black dashed lines) for θ (left) and λ (right).

proposal distributions, run the MCMC for 4,000 iterations and obtain an average acceptance rate of 32%. The top plot of Figure 1.2 provides the posterior draws (gray lines) and the progressive averages (black lines).

After a graphical inspection of the MCMC draws and of the progressive averages we can see the MCMC chain enter in the high probability region after about 1,000 iterations. Thus, we choose to discard an initial burn-in sample of 1,000 MCMC iterations and also to reduce dependence in the MCMC draws by removing one sample every two. Thus, the posterior approximation of all quantities of interest is based on 1,500 MCMC samples.

Bottom plots show the prior distributions (red solid lines) and kernel density estimates of the posterior distributions (black dashed lines) of λ and θ based on the MCMC samples. The comparison between prior and posterior distribution suggests that the prior has been revised and the data are informative about the value of the parameters. The Bayesian estimates of θ and λ are $\hat{\theta}_B = 2.35$ and $\hat{\lambda}_B = 0.24$, respectively. We find substantial evidence of over-dispersion which indicates the standard Poisson model is not well suited for this data. Figure 1.3 provides a comparison between the GP and the standard Poisson model and the empirical distribution of the cyber-attack frequency. The left plot suggests the GP (blue dots) can better capture than the Poisson (red dots) dispersion and fat tails of the empirical distribution. The right plot provides the 95% HPD region of the GP (blue area) and Poisson (red area) cumulative distributions.

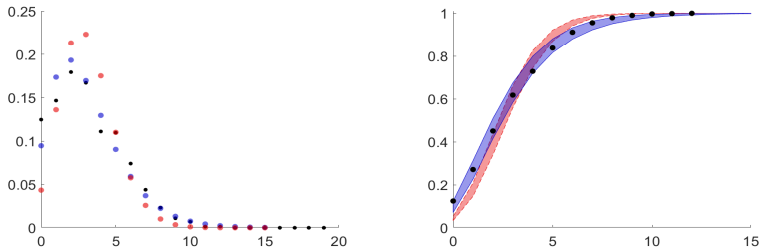


Figure 1.3: Left: empirical distribution of the data (black), generalized Poisson distribution (blue) at $\theta = \hat{\theta}_B = 2.35$ and $\lambda = \hat{\lambda}_B = 0.25$ and the standard Poisson distribution (red) at $\theta = \hat{\theta}_B = 3.14$. Right: empirical cumulative distribution of the data (black dots), the high posterior density region at the 95% level for the generalized Poisson distribution (blue) and the standard Poisson (red).

The empirical cumulative (black dots) is not entirely contained in the HPD of the Poisson model.

Top plots of Figure 1.4 provide the whole sample estimates (horizontal dashed lines), the sequential posterior mean (solid black lines) and sequential 95% HPD region (gray shaded area) over time on a rolling window of 120 observations. We find evidence of substantial temporal fluctuations in the parameters of the GP distribution.

More specifically, the sequential estimation of the mean $\mu^{(1)}$ (bottom-left plot of Figure 1.4) indicates that the expected number of cyber attacks increased from 2.62 in 2017 to 3.85 in 2018 (blue solid line, right axis). In the same plot the estimated coefficient of variation $C = \mu^{(2)}/\mu^{(1)}$ (red dashed line) increased from 1.15 in April 2017 to 2.03 in December 2018. The estimated values of $\mu^{(3)}$ and $\mu^{(4)}$ (bottom-right) indicate an increasing degree of asymmetry and tail heaviness in the distribution of the attack frequency.

1.4 Conclusion

Our Bayesian analysis of the cyber-attacks frequency provides evidence of over-dispersed data and of a better fitting of the generalised Poisson than the standard Poisson model. Our sequential analysis confirms the escalation of the cyber threats and the increased complexity of the phenomenon.

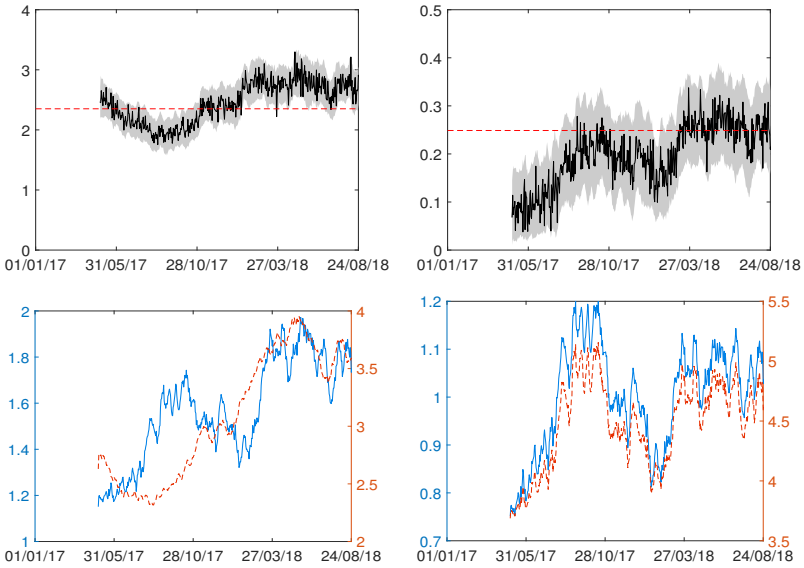


Figure 1.4: Sequential inference on a rolling window of 120 observations. Top: sequential estimates of θ (left plot, black solid lines) and λ (right plot, black solid lines) and their 95% HPD regions (gray areas). In both plots horizontal red dashed lines represent whole sample parameter estimates. Bottom: sequential coefficient of variation (left plot) and of the asymmetry $\mu^{(3)}$ (right plot, left scale) and kurtosis $\mu^{(4)}$ (right plot, right scale).

Bibliography

Cameron, A. C. and Trivedi, P. K. (1986). Econometric models based on count data. comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1(1):29–53.

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.

Consul, P. C. and Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics*, 15(4):791–799.

EIOPA (2018). Understanding cyber insurance: A structured dialogue with insurance groups. Available at `https://eiopa.europa.eu/Publications/Reports/EIOPAUnderstandingcyberinsurance.pdf`.

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

Hausman, J. A., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-rd relationship. Working Paper 17, National Bureau of Economic Research.

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.

Lee, L.-F. (1986). Specification test for poisson regression models. *International Economic Review*, 27(3):689–706.

Leslie, N. O., Harang, R. E., Knachel, L. P., and Kott, A. (2018). Statistical models for the number of successful cyber intrusions. *The Journal of Defense Modeling and Simulation*, 15(1):49–63.

Scollnik, D. P. (1998). On the analysis of the truncated generalized poisson distribution using a bayesian method. *ASTIN Bulletin: The Journal of the IAA*, 28(1):135–152.

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.

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.

Chapter 2

Generalised Poisson Difference Integer-Valued GARCH Process [1]

Abstract This chapter introduces a new stochastic process with values in the set $\mathbb{Z}$ of integers with sign. The increments of the process are Generalized 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, the stationary distribution of the process and its conditional and unconditional moments. We provide a Bayesian inference framework and an efficient posterior approximation procedure based on Markov Chain Monte Carlo. Numerical illustrations on simulated data show the effectiveness of the proposed inference. The applications to accidents data and cyber threats data show that the proposed model is well suited for capturing persistence in the conditional moments and in the over-dispersion feature of the data.
Keywords: Bayesian inference, Counts time series, Cyber risk, GARCH models.
MSC2010 subject classifications: 62G05, 62F15, 60G09, 60G57.

2.1 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)), sports (see Shahtahmassebi and Moyeed (2016)) and oceanography (see Cunha et al. (2018)). In this chapter, we build on Poisson models,

[1] In collaboration with Christian P. Robert (Université Paris-Dauphine, France and University of Warwick, UK). Appeared as working paper in https://arxiv.org/abs/2002.04470.

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. the Skellam distribution. Despite 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 underling 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.

Two main classes of time-varying parameter 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 chapter 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 allows 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 chapter, 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, 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 (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 an 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 chapter, 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 chapter 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 chapter, we develop a Bayesian inference procedure for the proposed GPD-INGARCH process and a Markov chain Monte Carlo (MCMC) procedure for posterior approximation. As argued in Ardia (2008) for non-Gaussian GARCH models, 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. Earlier example of Gibbs sampler using data augmentation is Tanner and Wong (1987). In count data, data augmentation has been implemented for logistic regression (e.g., see Fruehwirth-Schnatter and Frühwirth, 2007; Frühwirth-Schnatter and Frühwirth, 2010; Fussl et al., 2013), Poisson regression (Frühwirth-Schnatter and Wagner, 2006), Negative Binomial regression (Frühwirth-Schnatter et al., 2009) and for INARMA processes (Neal and Subba Rao, 2007).

We apply our model to a car accidents near the Schiphol airport dataset, previously studied in Brijs et al. (2008) and Andersson and Karlis (2014) with the main purpose of evaluating our model performance with respect to models previously used for this dataset. We introduced a second application to 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 chapter 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.

2.2 Generalized Poisson Difference INGARCH

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

$$P_x(\theta, \lambda) = \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-\theta - x\lambda} \qquad x = 0, 1, 2, \ldots \tag{2.1}$$

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:

$$P_z(\theta_1, \theta_2, \lambda) = e^{-\theta_1 - \theta_2 - z\lambda} \sum_{y=0}^{\infty} \frac{\theta_2(\theta_2 + y\lambda)^{y-1}}{y!} \frac{\theta_1(\theta_1 + (y+z)\lambda)^{y+z-1}}{(y+z)!} e^{-2y\lambda} \quad (2.2)$$

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 2.8.1 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 which will be used in this chapter. Appendix 2.8.2 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 an INGARCH model. A random variable Z follows a GPD if and only if its probability distribution is

$$P_z(\mu, \sigma^2, \lambda) = e^{-\sigma^2 - z\lambda} \sum_{s=\max(0,-z)}^{+\infty} \frac{1}{4} \frac{\sigma^4 + \mu^2}{s!(s+z)!} \left[\frac{\sigma^2 + \mu}{2} + (s+z)\lambda\right]^{s+z-1} \left[\frac{\sigma^2 - \mu}{2} + s\lambda\right]^{s-1} e^{-2\lambda s} \quad (2.3)$$

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

Remark 1. *The probability distribution in Eq. 2.3 is equivalent to the one in Eq. 2.2 up to the reparametrization $\mu = \theta_1 - \theta_2$ and $\sigma^2 = \theta_1 + \theta_2$. See Appendix 2.8.2 for a proof.*

The mean, variance, skewness and kurtosis of a GPD random variable can be obtained in closed-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:*

$$E(Z) = \frac{\mu}{1 - \lambda}, \qquad V(Z) = \frac{\sigma^2}{(1 - \lambda)^3} \quad (2.4)$$

and the Pearson skewness and kurtosis are:

$$S(Z) = \frac{\mu}{\sigma^3} \frac{(1 + 2\lambda)}{\sqrt{1 - \lambda}}, \qquad K(Z) = 3 + \frac{1 + 8\lambda + 6\lambda^2}{\sigma^2(1 - \lambda)} \quad (2.5)$$

(a) GPD(μ, σ^2, λ) distribution for $\sigma^2 = 10$ (b) GPD(μ, σ^2, λ) distribution for $\mu = 2$

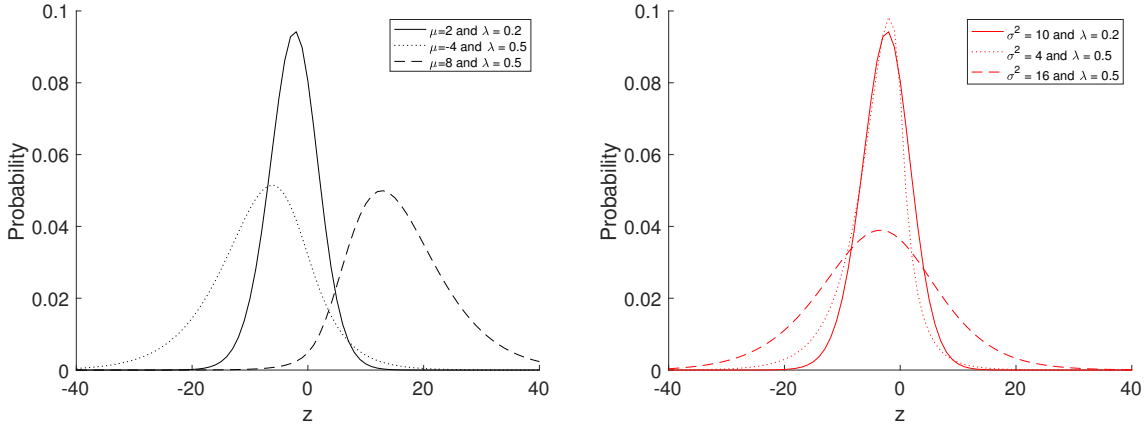


Figure 2.1: Generalized Poisson difference distribution GPD(μ, σ^2, λ) for some values of λ, μ and σ^2. The distribution with $\lambda = 0.2$, $\mu = 2$ and $\sigma^2 = 10$ (solid line) is taken as baseline in both panels.

See Appendix 2.8.2 for a proof.

Fig. 2.1 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 2.8.1 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

$$\frac{\tilde{\mu}_t}{1-\lambda} = \mu_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i} + \sum_{j=1}^{q} \beta_j \mu_{t-j} \tag{2.6}$$

where $\tilde{\mu}_{t-j} = \mu_{t-j}(1-\lambda)$, $\alpha_0 \in \mathbb{R}$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $i = 1, \ldots, p$, $p \geq 1$, $j = 1, \ldots, q$, $q \geq 0$. Another difference between our GPD-INGARCH and standard GARCH regards the parameters α_0. Since GPD distirbuted variables can take positive and negative values, modelling the location parameter allows for $\alpha_0 \in \mathbb{R}$. For $q = 0$ the model reduces to a GPD-INARCH(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:

$$\mu_t = \frac{\tilde{\mu}_t}{1-\lambda}, \quad \sigma_t^2 = \frac{\tilde{\sigma}_t^2}{(1-\lambda)^3} \tag{2.7}$$

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. 2.2, 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

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

where

$$\theta_{1t} = \frac{\sigma_t^2 + \mu_t}{2}, \quad \theta_{2t} = \frac{\sigma_t^2 - \mu_t}{2}. \tag{2.8}$$

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. 2.2) and high persistence, that is $\alpha_1 + \beta_1$ close to 1 (second column in Fig. 2.2). 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. 2.3 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 for a second-order stationary process to have an INGARCH representation. First define the two following polynomials: $D(B) = 1 - \beta_1 B - \ldots - \beta_q B^q$ and $G(B) = \alpha_1 B + \ldots + \alpha_p B^p$, 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

$$\mu_t = D^{-1}(B)(\alpha_0 + G(B)Z_t) = \alpha_0 D^{-1}(1) + H(B)Z_t \tag{2.9}$$

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. 2.6 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 2.8.2 □

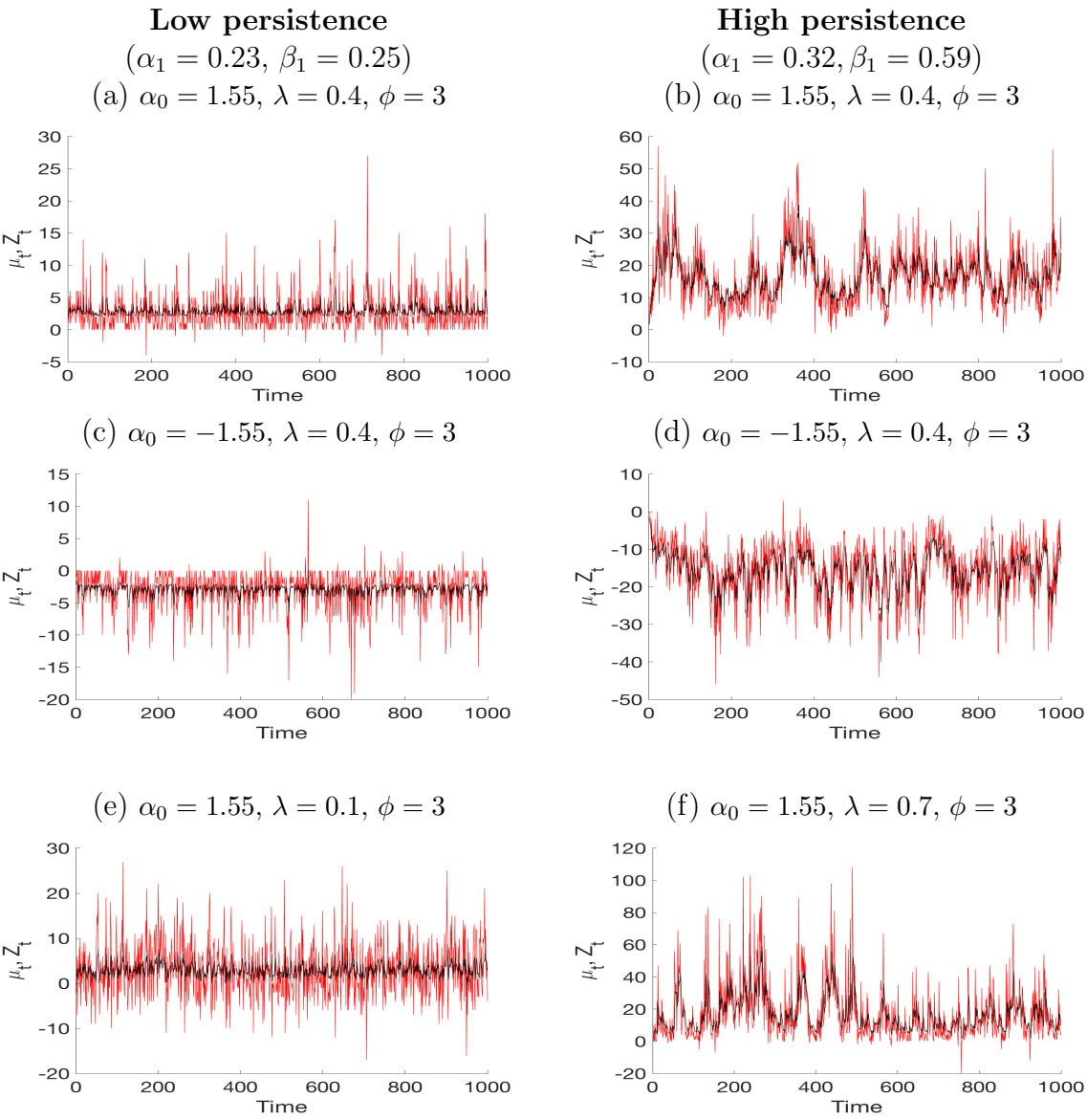


Figure 2.2: Simulated INGARCH(1, 1) paths for different values of the parameters α_0, α_1 and β_1. In Panels from (a) to (d) the effect of α_0 ($\alpha_0 > 0$ in the first line and $\alpha_0 < 0$ in the second line) with $\lambda = 0.4$ and $\phi = 3$. In Panels (e) and (f) the effect of lambda ($\lambda = 0.1$ left and $\lambda = 0.7$ right) in the two settings.

2.3 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 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.

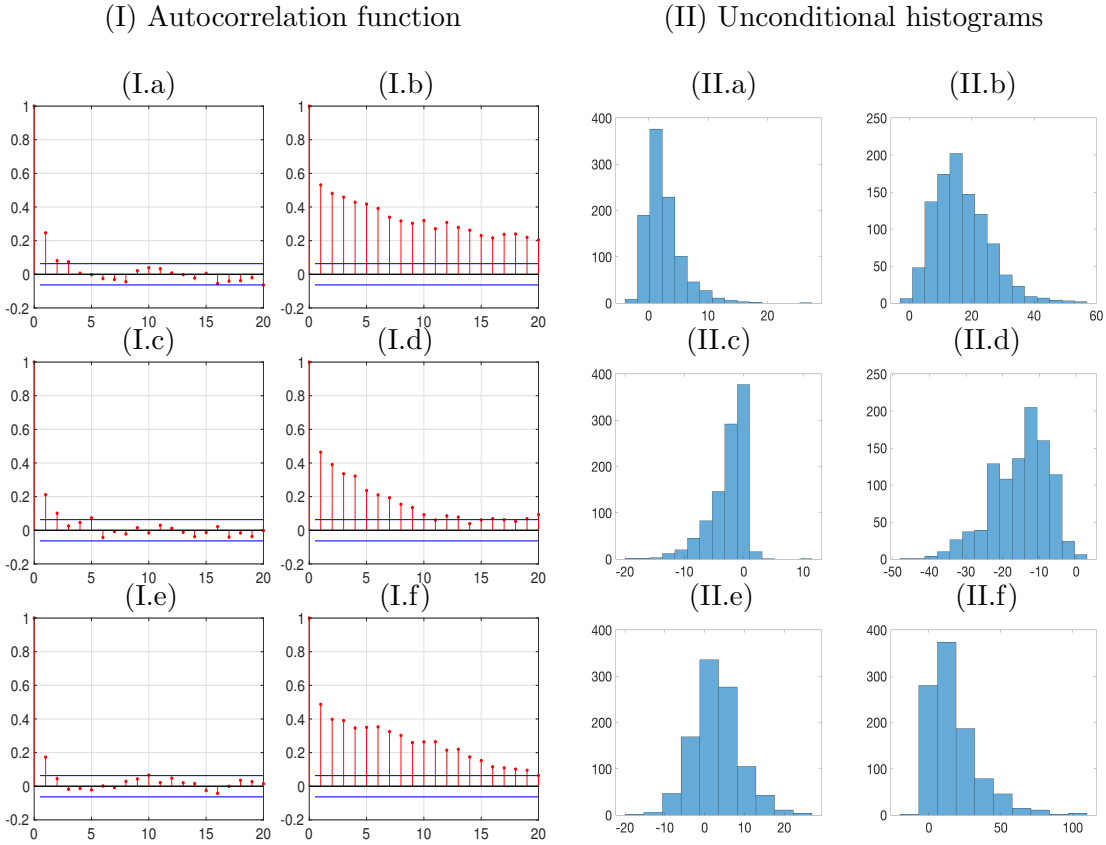


Figure 2.3: Autocorrelation functions (Panel I) and unconditional distributions (Panel II) of $\{Z_t\}_{t\in\mathcal{Z}}$ for the different cases presented in Fig. 2.2 (different columns in each panel).

2.3.1 Thinning representation

For the purpose of studying the properties of the GPD-INGARCH process, we show that it can be obtained as a limit of successive GPD-INAR approximations using the thinning operator. Let us define:

$$X_t^{(n)} = \begin{cases} 0, & n < 0 \\ (1-\lambda)U_{1t}, & n = 0 \\ (1-\lambda)U_{1t} + (1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{\frac{X_{t-i}^{(n-i)}}{(1-\lambda)}} V_{1t-i,i,j}, & n > 0 \end{cases} \tag{2.10}$$

and

$$Y_t^{(n)} = \begin{cases} 0, & n < 0 \\ (1-\lambda)U_{2t}, & n = 0 \\ (1-\lambda)U_{2t} + (1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{\frac{Y_{t-i}^{(n-i)}}{(1-\lambda)}} V_{2t-i,i,j}, & n > 0 \end{cases} \tag{2.11}$$

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(ρ,θ/λ,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($\rho\lambda$,θ) distribution.*

Proof. See Alzaid and Al-Osh (1993). □

Both $X_t^{(n)}$ and $Y_t^{(n)}$ in Eq. 2.10 and 2.11 admit the representation

$$X_t^{(n)} = (1-\lambda)U_{1t} + (1-\lambda)\sum_{i=1}^{n} \varphi_{1i}^{(t-i)} \circ \left(\frac{X_{t-i}^{(n-i)}}{1-\lambda}\right),\ n > 0 \tag{2.12}$$

and

$$Y_t^{(n)} = (1-\lambda)U_{2t} + (1-\lambda)\sum_{i=1}^{n} \varphi_{2i}^{(t-i)} \circ \left(\frac{Y_{t-i}^{(n-i)}}{1-\lambda}\right),\ n > 0 \tag{2.13}$$

where $\varphi \circ X$ is the quasi-binomial thinning operation. See Appendix 2.8.1 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:

$$\rho \diamond Z | Z \overset{\mathrm{d}}{=} (\rho_{\theta_1,\lambda} \circ X) - (\rho_{\theta_2,\lambda} \circ Y) | (X - Y) \tag{2.14}$$

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 "$A \overset{\mathrm{d}}{=} B$" 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 2.8.1 for further details. Using the new operator as defined in Eq. 2.14, we can represent $Z_t^{(n)}$ as follows.

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

$$Z_t^{(n)} = (1-\lambda)U_t + (1-\lambda)^2 \sum_{i=1}^{n} \varphi_i^{(t-i)} \diamond \left(\frac{Z_{t-i}^{(n-i)}}{1-\lambda}\right),\ n > 0 \tag{2.15}$$

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 2.8.2 □

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:

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

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

$$
\begin{aligned}
\mu_n &= (1-\lambda)E[U_t] + (1-\lambda)^2 \sum_{i=1}^{n} E\left[\varphi_i^{(t-i)} \diamond \left(\frac{Z_{t-i}^{(n-i)}}{1-\lambda}\right)\right] \\
&= \psi_0 + \sum_{j=1}^{\infty} \psi_j \mu_{n-j} = D^{-1}(B)\alpha_0 + H(B)\mu_n
\end{aligned}
\tag{2.16}
$$

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 2.8.2. □

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

Proof. See Appendix 2.8.2. □

2.3.2 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 $\mathbf{W} = (W_1, \ldots, W_k)$

$$g_{\mathbf{W}}(\mathbf{t}) = E\left[\prod_{i=1}^{k} t_i^{W_i}\right] = \sum_{\mathbf{W}\in\mathbb{N}^k} p(\mathbf{w}) \prod_{i=1}^{k} t_i^{W_i} \tag{2.17}$$

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

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

$$g_{\mathbf{Z}_{1\ldots k}}(\mathbf{t}) = g_{\mathbf{X}_{1\ldots k}}(\mathbf{t}) g_{\mathbf{Y}_{1\ldots k}}(\mathbf{t^{-1}}) \tag{2.18}$$

Proof. See Appendix 2.8.2 □

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.*

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 2.8.2. □

Since $\{Z_t\}_{t\in\mathbb{Z}}$ is the almost sure limit of $\{Z_t^{(n)}\}$ one can find the conditional distribution of $\{Z_t\}_{t\in\mathbb{Z}}$ using the properties of the thinning representation.

Proposition 10. *Let $\mathcal{F}_{t-1} = \sigma(\{Z_u\}_{u\leq t-1})$, for $t \in \mathbb{Z}$, the conditional law of $\{Z_t^{(n)}\}_{t\in\mathbb{Z}}$ given $\mathcal{F}_{t-1}$ converges to a $GPD(\tilde{\mu}_t, \tilde{\sigma}_t^2, \lambda)$.*

2.3.3 Moments of the GPD-INGARCH

The conditional mean and variance of the process Z_t are

$$E(Z_t|\mathcal{F}_{t-1}) = \frac{\tilde{\mu}_t}{1-\lambda} = \mu_t, \quad V(Z_t|\mathcal{F}_{t-1}) = \frac{\tilde{\sigma}_t^2}{1-\lambda} = \phi^3\tilde{\sigma}_t^2 \tag{2.19}$$

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

The unconditional mean and variance of the process are

$$E(Z_t) = \mu_t = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}$$

$$\begin{aligned} V(Z_t) &= E\left[V\left(Z_t|\mathcal{F}_{t-1}\right)\right] + V\left[E\left(Z_t|\mathcal{F}_{t-1}\right)\right] \\ &= E(\phi^3\tilde{\sigma}_t^2) + V(\mu_t) \\ &= \phi^3 E(\tilde{\sigma}_t^2) + V(\mu_t) \end{aligned} \tag{2.20}$$

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. 2.6 with $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. 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

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

$$\gamma_\mu(k) = \sum_{i=1}^{min(k,p)} \alpha_i \gamma_\mu(|k-i|) + \sum_{i=k+1}^{p} \alpha_i \gamma_Z(i-k) + \sum_{j=1}^{q} \beta_j \gamma_\mu(|k-j|), \quad k \geq 0. \tag{2.21}$$

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 2.8.2.

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

$$\mu_t = \alpha_0 + \alpha_1 Z_{t-1} \tag{2.22}$$

then the linear equations in Eq. 2.21, become

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

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

$$\gamma_Z(k) = \begin{cases} \alpha_1^{k-1}\gamma_Z(1), & \text{for } k \geq 2 \\ \alpha_1[\phi^3 E(\tilde{\sigma}_t^2)] + \alpha_1 V(\mu_t), & \text{for } k = 1 \end{cases} \tag{2.23}$$

$$\gamma_\mu(k) = \begin{cases} \alpha_1^k V(\mu_t), & \text{for } k \geq 1 \\ \alpha_1^2[\phi^3 E(\tilde{\sigma}_t^2)] + \alpha_1^2 V(\mu_t), & \text{for } k = 0 \end{cases} \tag{2.24}$$

Therefore, the variance of μ_t is

$$V(\mu_t) = \frac{\alpha_1^2[\phi^3 E(\tilde{\sigma}_t^2)]}{1-\alpha_1^2} \tag{2.25}$$

and the variance of Z_t is

$$V(Z_t) = \frac{\phi^3 E(\tilde{\sigma}_t^2)}{1-\alpha_1^2} \tag{2.26}$$

where $\phi = \frac{1}{1-\lambda}$.
Lastly, the autocorrelations are

$$\rho_\mu(k) = \alpha_1^k \tag{2.27}$$

$$\rho_Z(k) = \alpha_1^k \tag{2.28}$$

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

$$\mu_t = \alpha_0 + \alpha_1 Z_{t-1} + \beta_1 \mu_{t-1} \tag{2.29}$$

From Eq. 2.21,

$$\gamma_Z(k) = \begin{cases} (\alpha_1+\beta_1)^{k-1}\gamma_Z(1), & \text{for } k \geq 2 \\ \alpha_1[\phi^3 E(\tilde{\sigma}_t^2)] + (\alpha_1+\beta_1)V(\mu_t), & \text{for } k = 1 \end{cases} \tag{2.30}$$

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

$$\gamma_\mu(k) = \begin{cases} (\alpha_1+\beta_1)^k V(\mu_t), & \text{for } k \geq 1 \\ \alpha_1^2[\phi^3 E(\tilde{\sigma}_t^2)] + (\alpha_1+\beta_1)^2 V(\mu_t), & \text{for } k = 0 \end{cases} \tag{2.31}$$

where the second equation in Eq. 2.31 is equal to $V(\mu_t)$. From this latter equation, we can derive the expression for $V(\mu_t)$

$$V(\mu_t) = \frac{\alpha_1^2[\phi^3 E(\tilde{\sigma}_t^2)]}{1-(\alpha_1+\beta_1)^2} \tag{2.32}$$

Combining Eq. 2.20 and 2.32, we can derive a close expression for the variance of Z_t:

$$V(Z_t) = \frac{\phi^3 E(\tilde{\sigma}_t^2)[1-(\alpha_1+\beta_1)^2+\alpha_1^2]}{1-(\alpha_1+\beta_1)^2} \tag{2.33}$$

The autocorrelations are given by

$$\rho_\mu(k) = (\alpha_1+\beta_1)^k \tag{2.34}$$

$$\rho_Z(k) = (\alpha_1+\beta_1)^{k-1}\,\frac{\alpha_1[1-\beta_1(\alpha_1+\beta_1)]}{1-(\alpha_1+\beta_1)^2+\alpha_1^2} \tag{2.35}$$

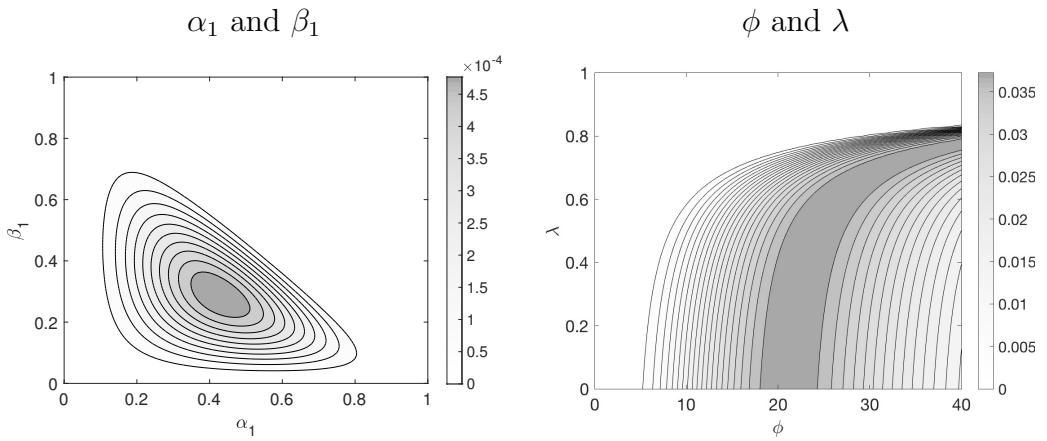


Figure 2.4: Contour lines of the log-prior density function for α_1 and β_1 (left) and ϕ and λ (right).

2.4 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.

2.4.1 Prior assumption

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

$$\pi(\boldsymbol{\varphi}) = \frac{\Gamma\left(\sum_{i=0}^{d} c_i\right)}{\prod_{i=0}^{d} \Gamma(c_i)} \prod_{i=1}^{d} \varphi_i^{c_i - 1} \left(1 - \sum_{i=1}^{d} \varphi_i\right)^{(c_0 - 1)} \tag{2.36}$$

where $\varphi_i \geq 0$ and $\sum_{i=1}^{d} \varphi_i \leq 1$ (Wong, 1998). Panel (a) in Fig. 2.4 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 $c_2 = 3$. For λ and ϕ we assume a joint prior distribution with uniform marginal prior $\lambda \sim \mathcal{U}_{[0,1]}$ and shifted gamma conditional prior $\phi \sim \mathcal{G}a^*(a, b, c)$, with density function:

$$\pi(\phi) = \frac{b^a}{\Gamma(a)} (\phi - c)^{(a-1)} e^{-b(\phi - c)} \quad \text{for } \phi > c \tag{2.37}$$

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 $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\varphi})\pi(\alpha_0)\pi(\lambda)\pi(\phi)$. Since we found high autocorrelation in the estimation of α_0, in the simulation study and in the real data application we fix the value of α_0, to a value close to the unconditional mean, to improve the efficiency of the MCMC algorithm.

2.4.2 Data augmentation

Denote the probability distribution of Z_t with

$$f_t(Z_t = z|\boldsymbol{\theta}) = e^{-\sigma_t^2 - z\lambda} \sum_{s=\underline{s}}^{+\infty} \frac{1}{4} \frac{\sigma_t^4 + \mu_t^2}{s!(s+z)!} \left[\frac{\sigma_t^2 + \mu_t}{2} + (s+z)\lambda\right]^{s+z-1} \left[\frac{\sigma_t^2 - \mu_t}{2} + s\lambda\right]^{s-1} e^{-2\lambda s} \tag{2.38}$$

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

$$\pi(\boldsymbol{\theta}|Z_{1:T}) \propto \prod_{t=1}^{T} f_t(Z_t|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{2.39}$$

is not analytically tractable we apply Markov Chain Monte Carlo (MCMC) for posterior approximation in combination with a data-augmentation approach (Tanner and Wong, 1987). Alternative approaches can be nested Laplace approximation (e.g., see Gómez-Rubio and Rue, 2018; Rue et al., 2009) and particle Gibbs sampling (Andrieu et al., 2010). Our model is still tractable, hence we don't need this alternative methods. See Robert and Casella (2013) for an introduction to MCMC. As in Karlis and Ntzoufras (2006), we exploit the stochastic representation in Eq. 2.8 and introduce two GP latent variables X_t and Y_t with pmfs

$$f_t(X_t = x|\theta_{1t}, \lambda) = \frac{\theta_{1t}(\theta_{1t} + \lambda x)^{x-1}}{x!} e^{(-\theta_{1t} - \lambda x)} \tag{2.40}$$

$$f_t(Y_t = y|\theta_{2t}, \lambda)) = \frac{\theta_{2t}(\theta_{2t} + \lambda y)^{y-1}}{y!} e^{(-\theta_{2t} - \lambda y)}, \tag{2.41}$$

where the subscript index t to the pmfs indicates the dependence of the two latent variables on the past values of the observed process Z_{t-k}, $k \geq 1$, through the dynamic parameters θ_{1t} and θ_{2t}. Note that, the Z_t are conditionally independent given the two latent variables. We summarize our Bayesian model in the Directed Acyclic Graph (DAG) representation of Fig. 2.5.

Let $Z_{1:T} = (Z_1, \ldots, Z_T)$, $X_{1:T} = (X_1, \ldots, X_T)$ and $Y_{1:T} = (Y_1, \ldots, Y_T)$. Following the conditional independence of the observables, the complete-data likelihood factorizes as follows

$$\begin{aligned} f(Z_{1:T}, X_{1:T}Y_{1:T}|\boldsymbol{\theta}) &= \prod_{t=1}^{T} f(Z_t|X_t, Y_t, \boldsymbol{\theta}) f_t(X_t, Y_t|\boldsymbol{\theta}) \\ &= \prod_{t=1}^{T} \delta(Z_t - X_t + Y_t) f_t(X_t|\boldsymbol{\theta}) f_t(X_t - Z_t|\boldsymbol{\theta}). \end{aligned} \tag{2.42}$$

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

$$\pi(X_{1:T}, Y_{1:T}, \boldsymbol{\theta}|Z_{1:T}) \propto f(Z_{1:T}, X_{1:T}Y_{1:T}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{2.43}$$

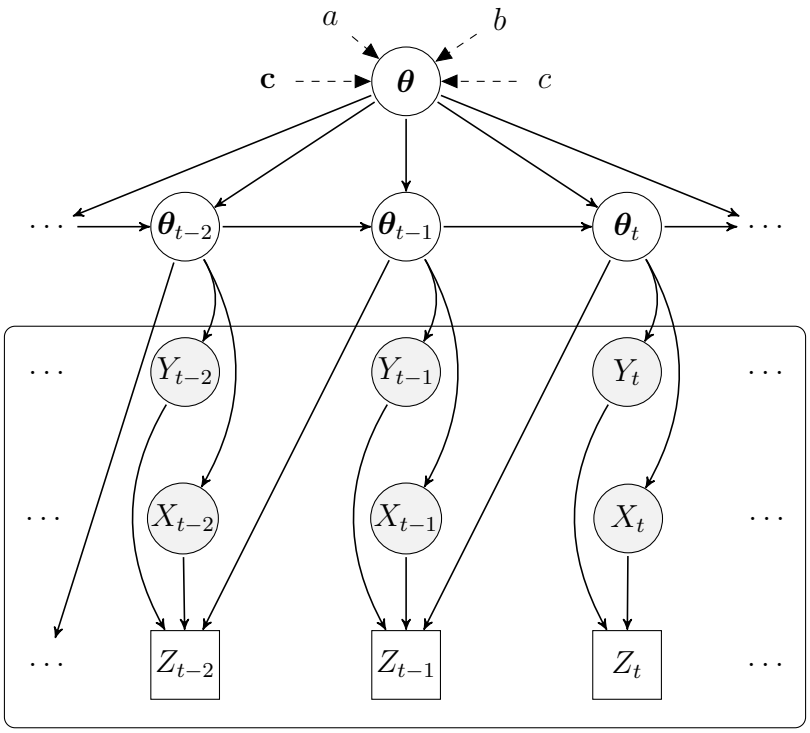


Figure 2.5: DAG of the Bayesian GPD-INGARCH(1,1) model. It exhibits the hierarchical structure of the observations Z_t (boxes), the latent variables X_t and Y_t (gray circles), the time-varying and the static parameters, $\boldsymbol{\theta}_t = (\theta_{1t}, \theta_{2t})$ and $\boldsymbol{\theta} = (\boldsymbol{\varphi}, \alpha_0, \lambda, \phi)$, respectively (white circles) and the hyperparameters $\mathbf{c}$, a, b, c. The directed arrows show the causal dependence structure of the model.

2.4.3 MCMC sampler

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

where $\boldsymbol{\theta}_{-\eta}$ indicates the collection of parameters excluding the element $\boldsymbol{\eta}$.

The full conditional for the latent variables is

$$(X_t, Y_t) \sim f(Z_t | X_t, Y_t, \boldsymbol{\theta}) f_t(X_t, Y_t | Z_t, \boldsymbol{\theta}). \tag{2.44}$$

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

$$\varrho = \min\left\{1, \frac{f_t(X_t^*|\theta_{1t},\lambda) f_t(X_t^* - Z_t|\theta_{2t},\lambda)}{f_t(X_t^{(j-1)}|\theta_{1t},\lambda) f_t(X_t^{(j-1)} - Z_t|\theta_{2t},\lambda)} \frac{q(X_t^{(j-1)})}{q(X_t^*)}\right\} \tag{2.45}$$

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

$$\boldsymbol{\varphi} \sim \pi(\boldsymbol{\varphi}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\varphi}) \propto \pi(\boldsymbol{\varphi}) \prod_{t=1}^{T} f_t(X_t, Y_t|\boldsymbol{\theta}). \tag{2.46}$$

We consider a MH with Dirichlet independent proposal distribution

$$\boldsymbol{\varphi}^* \sim Dir(\mathbf{c}^*) \tag{2.47}$$

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

$$\varrho = \min\left\{1, \frac{\pi(\boldsymbol{\varphi}^*|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\varphi})}{\pi(\boldsymbol{\varphi}^{j-1}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\varphi})}\right\}. \tag{2.48}$$

The full conditional distribution of ϕ is

$$\pi(\phi|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi}) \propto \pi(\phi) \prod_{t=1}^{T} f_t(X_t, Y_t|\boldsymbol{\theta}). \tag{2.49}$$

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

$$\zeta^* \sim N(\zeta_{j-1}, \gamma^2) \tag{2.50}$$

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

$$\varrho = \min\left\{1, \frac{\pi(\phi^*|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi}) \exp(\zeta^*)}{\pi(\phi_{j-1})|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\phi}) \exp(\zeta_{j-1})}\right\} \tag{2.51}$$

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

The full conditional distribution of λ is

$$\pi(\lambda|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda}) \propto \pi(\lambda) \prod_{t=1}^{T} f_t(X_t, Y_t|\boldsymbol{\theta}). \tag{2.52}$$

We consider a MH step with Beta random walk proposal

$$\lambda^* \sim Be(s\lambda^{(j-1)}, s(1-\lambda^{(j-1)})) \tag{2.53}$$

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

$$\varrho = \min\left\{1, \frac{\pi(\lambda^*|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda})Be(s\lambda^*, s(1-\lambda^*))}{\pi(\lambda^{(j-1)}|Z_{1:T}, Y_{1:T}, X_{1:T}, \boldsymbol{\theta}_{-\lambda})Be(s\lambda^{(j-1)}, s(1-\lambda^{(j-1)}))}\right\}. \tag{2.54}$$

To have an acceptance rate in the range $[0.15, 0.5]$, the values for the hyper-parameters of the proposal are: $= (3, 3, 4)$, $\gamma \sim N(0, 0.0001)$, $s = 100$ while the parameter ν is set to be the proportion of positive values in the series.

2.4.4 Forecasting

Given the samples $Z_1, \ldots, Z_T$, the Gibbs sampler can be used to approximate point and distribution forecasts for the variable of interest Z_{T+h}, $h = 1, \ldots, H$ where H is the forecasting horizon. At the j-th MCMC iteration we draw $Z^{(j)}_{T+h}$ from the conditional distribution given past observations and the parameter draw $\boldsymbol{\theta}^{(j)}$

$$Z^{(j)}_{T+h}|Z^{(j)}_{T+h-1}, \mathcal{F}_T, \boldsymbol{\theta}^{(j)} \sim GPD\left(\tilde{\mu}^{(j)}_{T+h}, \tilde{\sigma}^{2\,(j)}_{T+h}, \lambda^{(j)}\right) \tag{2.55}$$

$h = 1, \ldots, H$, where $\tilde{\mu}^{(j)}_{T+h} = \mu^{(j)}_{T+h}\left(1-\lambda^{(j)}\right)$, $\tilde{\sigma}^{2\,(j)}_{T+h} = \sigma^{2\,(j)}_{T+h}\left(1-\lambda^{(j)}\right)^2$ and

$$\mu^{(j)}_{T+h} = \begin{cases} \alpha^{(j)}_0 + \alpha^{(j)}_1 Z_T + \beta^{(j)}_1 \mu^{(j)}_T, & \text{for } h = 1 \\ \alpha^{(j)}_0 + \alpha^{(j)}_1 Z^{(j)}_{T+h-1} + \beta^{(j)}_1 \mu^{(j)}_{T+h-1}, & \text{for } h = 2, \ldots, H \end{cases} \tag{2.56}$$

$$\sigma^{2\,(j)}_{T+h} = |\mu^{(j)}_{T+h}|\phi^{(j)} \tag{2.57}$$

with $j = 1, \ldots, J$ indicating the MCMC draws. The point forecast $\mathbb{E}(Z_{T+h}|\mathcal{F}_T)$ can be approximated as follows

$$\widehat{\mathbb{E}(Z_{T+h}|\mathcal{F}_T)} = \frac{1}{J}\sum_{j=1}^{J} Z^{(j)}_{T+h} \tag{2.58}$$

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

For example, in some applications the process $\{Z_t\}_{t\in\mathbb{Z}}$ represents the time increments of the process $\{V_t\}_{t\in\mathbb{Z}}$ that is $Z_t = V_t - V_{t-1}$. Random samples for $V_{T+h} = V_{T+h-1} + Z_{T+h}$ can be easily obtained from the recursion

$$V^{(j)}_{T+h} = V^{(j)}_{T+h-1} + Z^{(j)}_{T+h}, \quad h = 1, \ldots, H \tag{2.59}$$

with initial value $V^{(j)}_T$ equal to the sample V_T available at time T, where the GPD increments $Z^{(j)}_{T+h}$ are sampled from $GPD\left(\tilde{\mu}^{(j)}_{T+h}, \tilde{\sigma}^{2\,(j)}_{T+h}, \lambda^{(j)}\right)$ under the constrain: $Z^{(j)}_{T+h} \geq V^{(j)}_{T+h-1}$. The point forecast $\mathbb{E}(V_{T+h}|\mathcal{F}_T)$ can be approximated as follows

$$\widehat{\mathbb{E}(V_{T+h}|\mathcal{F}_T)} = \frac{1}{J}\sum_{j=1}^{J} V^{(j)}_{T+h}. \tag{2.60}$$

2.5 Simulation study

The purpose of our simulation exercises is threefold: first we test for the correct implementation of the MCMC, algorithm presented in Section 2.4, second we study the efficiency of the MCMC and finally we show the effect of misspecification of the overdispersion parameter λ on the estimates of the GARCH coefficient.

We apply the Geweke (2004) procedure to test the correct implementation of the Gibbs sampler. The test used 2000 MCMC samples and the first three moments as test functions for each parameter. Our Gibbs sampler passes the tests. Background material on the Geweke's test and the value of the statistics are reported in Appendix 2.8.3.

As regards the efficiency of the MCMC we evaluated the Geweke (1992) convergence diagnostic measure (CD), the inefficiency factor (INEFF)[2] 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 to remove dependence on initial conditions, and finally apply a thinning procedure with a factor of 250, to reduce the dependence between consecutive draws. For each setting we provide in Appendix 2.8.3 the output of one of the experiments. Since we obtain better results than previous models, we set $p = q = 1$ and in the following for ease of notation we use $\alpha_1 = \alpha$ and $\beta_1 = \beta$. We leave the study for models with $p + q > 2$ for future works.

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 4.1 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

[2]The inefficiency factor is defined as

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

where $\rho(k)$ is the sample autocorrelation at lag k for the parameter of interest 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	-28.69	0.78	0.93	-6.27	2.40
	(0.11)	(0.14)	(0.10)	(0.04)	(0.06)	(0.05)
CD_{AT}	5.72	-13.18	0.2	0.74	-3.84	1.17
	(0.23)	(0.23)	(0.23)	(0.13)	(0.15)	(0.11)

Table 2.1: 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.

MCMC after 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).

Finally, we study the estimation bias when the model is not correctly specified. Since the GARCH dynamics in μ_t is related to the over-dispersion λ and location $\tilde{\mu}_t$ (see equation 2.19) one can expect some estimation bias when λ is not correctly specified. In our experiments, when a PD-INGARCH is estimated on data generated from a GPD-INGARCH, the GARCH parameters are biased. For illustrative purposes we report in Figure 2.6 the results of some of our experiments. The estimation bias for α and β is 0.11 and -0.09, respectively in the low persistence case and 0.1 and 0.02, respectively in the high persistence.

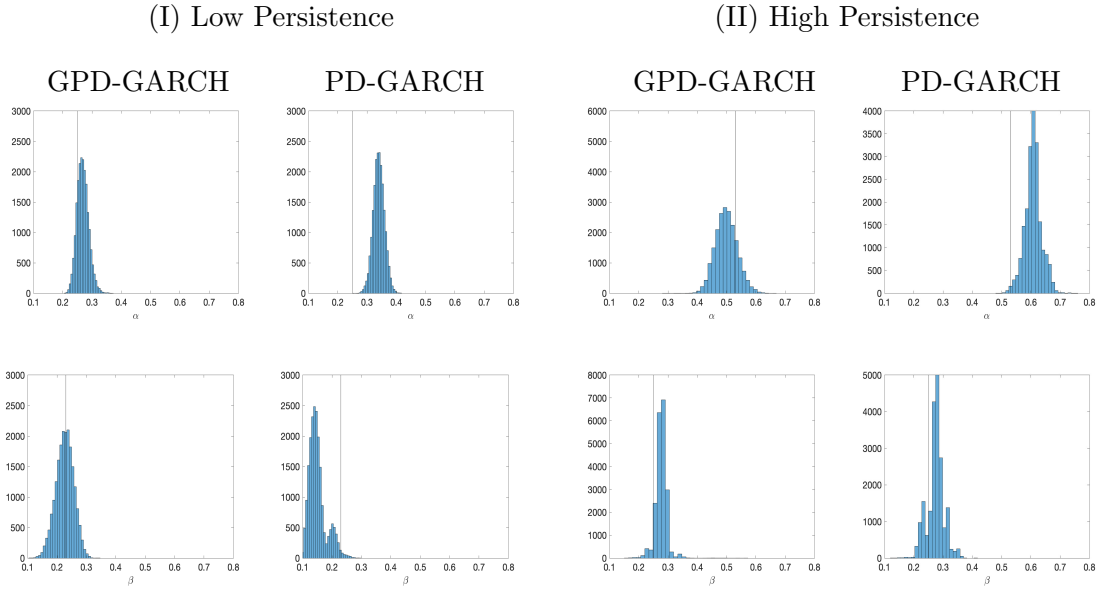


Figure 2.6: Posterior histograms for α (first row) and β (second row) in the low persistence case (left panel) and high persistence case (right panel) when we fit the correct model $GPD-INGARCH$ and the misspecified model $PD-INGARCH$.

2.6 Real data examples

2.6.1 Accident data

Data in this application are the number of daily car accidents near Schiphol airport in The Netherlands during 2001 (Fig. 2.7) sampled at daily frequency. This set of 365 observations have been previously considered in Brijs et al. (2008) and Andersson and Karlis (2014). Following the results of the Augmented Dickey-Fueller and Phillips-Perron tests for unit roots, the time series of accident counts is non-stationary, whereas the first differences do not exhibit unit root (see Tab. 2.6 in Appendix 2.8.4). We applied our Bayesian estimation procedure, as described in Section 2.4.

For each parameters, in Fig. 2.8 are presented the histograms for the Gibbs draws. Table 2.2 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 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 heteroskedasticy and persistence by testing some restrictions of the INGARCH(1,1) (models from $\mathcal{M}_2$ to $\mathcal{M}_4$ in Table 2.2).

Bayesian inference compares models via the so-called Bayes factor, which is the ratio of normalizing constants of the posterior distributions of two different

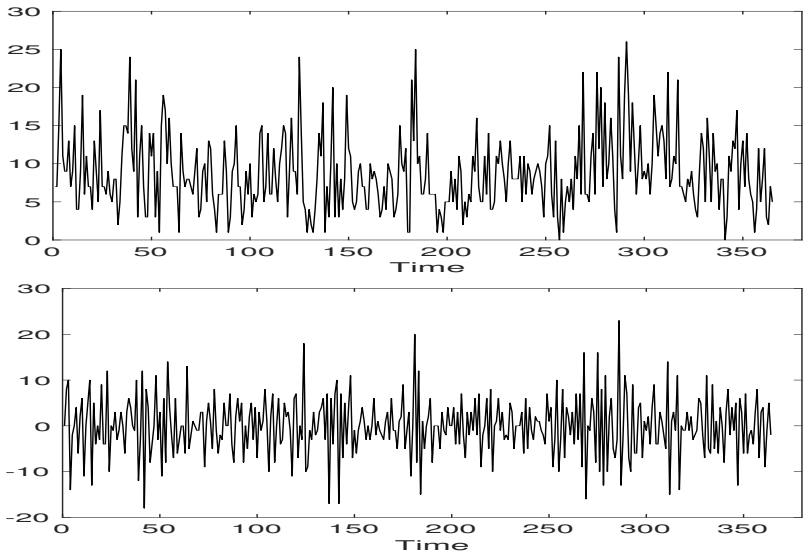


Figure 2.7: Frequency (top) and month-on-month changes (bottom) of the accidents at the Schiphol airport in The Netherlands in 2001.

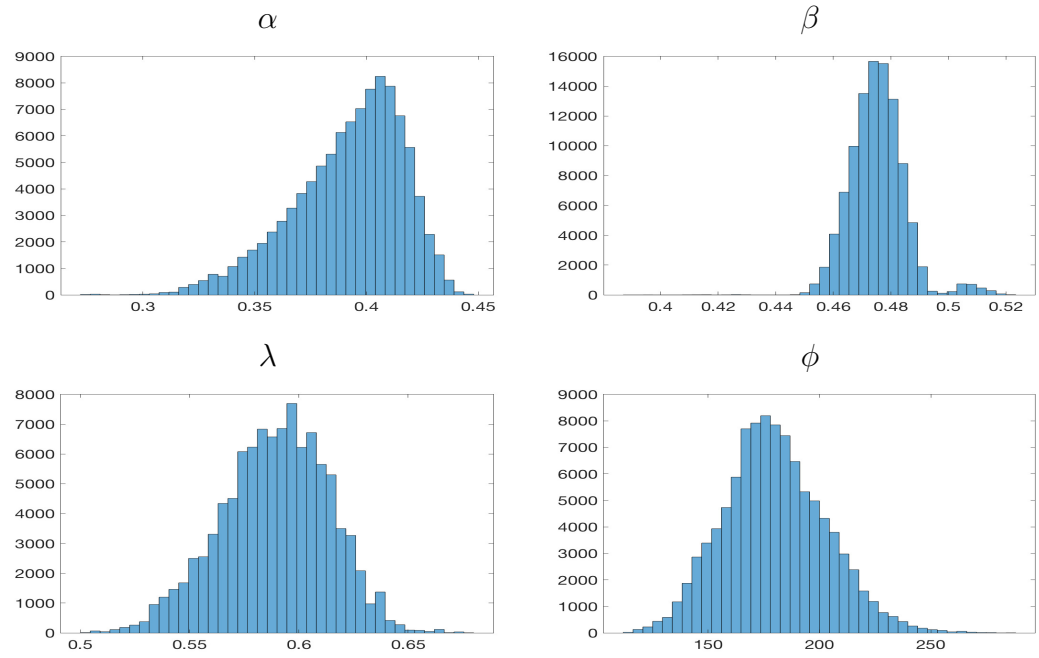


Figure 2.8: Histograms of the MCMC draws for the parameters of the Schiphol's accident data of Fig. 3.2.

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

models (see Cameron et al. (2014) for a review). MCMC methods allow for generating samples from the posterior distributions which can be used to estimate the ratio of normalizing constants.

In this chapter 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

$$p_j(x, \eta) = \frac{h_j(x)\exp(\eta_j)}{h_1(x)\exp(\eta_1) + h_2(x)\exp(\eta_2)}, \ j = 1, 2 \tag{2.61}$$

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

$$\ell_n(\kappa) = \sum_{i=1}^{n} \log p_1(X_{i1}, \eta_1) + \sum_{i=1}^{n} \log p_2(X_{i2}, \eta_2) \tag{2.62}$$

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 Table 2.2.

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

Table 2.3: Logarithmic Bayes Factor, BF($\mathcal{M}_i$,$\mathcal{M}_j$), of the model $\mathcal{M}_i$ (rows) against model $\mathcal{M}_j$ (columns), with $i < j$. Where $\mathcal{M}_1$ is the GPD-INGARCH(1,1), $\mathcal{M}_2$ is the PD-INGARCH(1,1) with $\lambda = 0$, $\mathcal{M}_3$ is the GPD-INARCH(1,0) and $\mathcal{M}_4$ is the PD-INARCH(1,0) with $\lambda = 0$. Numbers in parenthesis are standard deviations of the estimated Bayes factors.

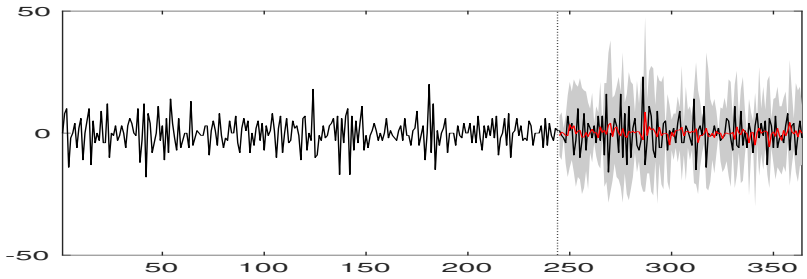


Figure 2.9: Changes in the number of car accidents and their sequential one-step-ahead forecast (red line) and 95% HPD region (gray area)

We performed six reverse logistic regressions, in which we compare our models pairwise. The approximated logarithmic Bayes factors $BF(\mathcal{M}_i, \mathcal{M}_j)$ are given in Table 2.3. 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$(1, 0)$.

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 car accidents. We account for parameter uncertainty and approximate the predictive distribution by MCMC.

2.6.2 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 chapter we

use the classification provided in the Passeri (2019) dataset. Passeri (2019) 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.

Fig. 2.10 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 2.2.

We applied our estimation procedure, as described in Section 2.4. 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. 2.11 are presented the histograms for the Gibbs draws for each parameters.

BF($\mathcal{M}_i$,$\mathcal{M}_j$)	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$\mathcal{M}_1$	4.72e+03	-558.34	-229.88
	(1.068)	(-0.115)	(-0.028)
$\mathcal{M}_2$		102.567	102.567
		(2.737e-15)	(8.024e-16)
$\mathcal{M}_3$			102.567
			(4.447e-17)

Table 2.4: Logarithmic Bayes Factor, BF($\mathcal{M}_i$,$\mathcal{M}_j$), of the model $\mathcal{M}_i$ (rows) against model $\mathcal{M}_j$ (columns), with $i < j$. Where $\mathcal{M}_1$ is the GPD-INGARCH(1,1), $\mathcal{M}_2$ is the PD-INGARCH(1,1) with $\lambda = 0$, $\mathcal{M}_3$ is the GPD-INARCH(1,0) and $\mathcal{M}_4$ is the PD-INARCH(1,0) with $\lambda = 0$. Numbers in parenthesis are standard deviations of the estimated Bayes factors.

Fig. 2.11 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 suggests the presence of over-dispersion in the data.

As in the previous application, we performed six reverse logistic regressions, in which we compare our models pairwise. The approximated logarithmic Bayes factors $BF(\mathcal{M}_i, \mathcal{M}_j)$ are given in Table 2.4. 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$(1, 0)$.

Given the importance of forecasting cyber-attacks, in this section we present the

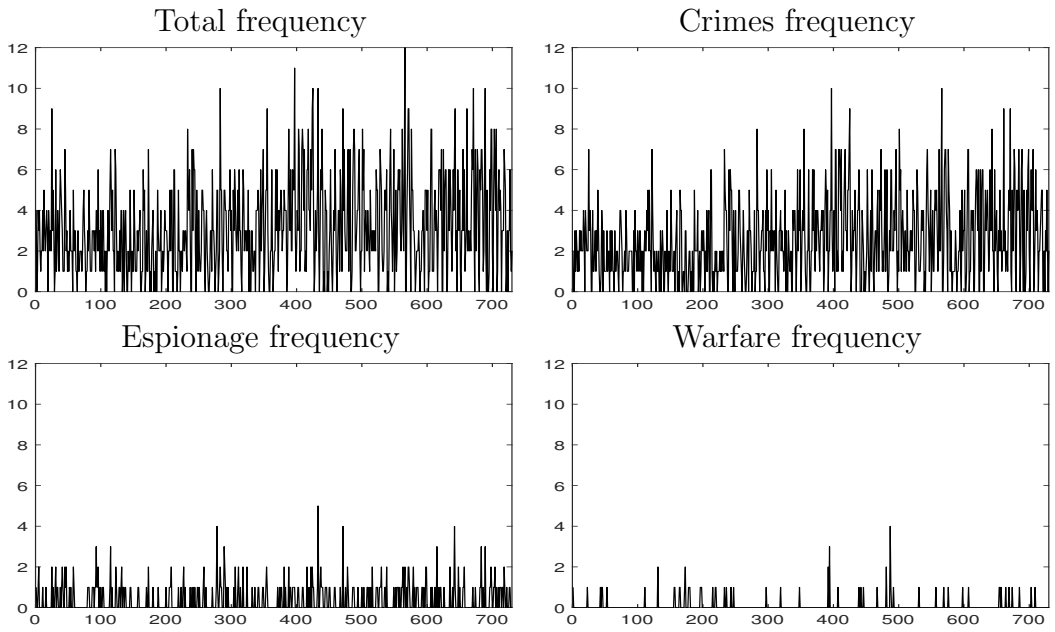


Figure 2.10: Daily cyber-threats counts between 1st January 2017 and 31st December 2018.

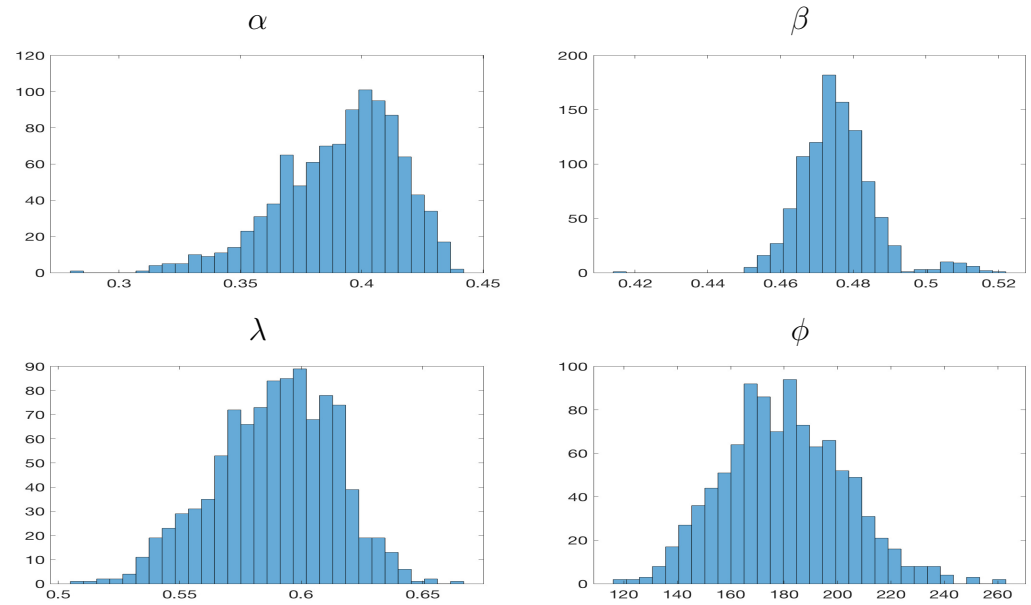


Figure 2.11: Histograms of the MCMC draws for the parameters.

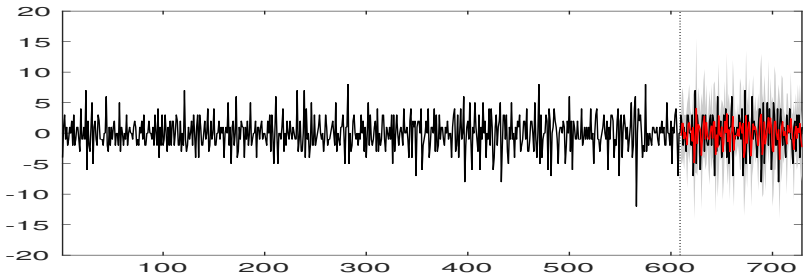


Figure 2.12: Changes in the number of cyber threats and their sequential one-step-ahead forecast (red line) and 95% HPD region (gray area) between 1st November 2018 (vertical dashed line) and 31st December 2018.

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.

2.7 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.

Bibliography

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.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342.

Ardia, D. (2008). Bayesian estimation of a Markovswitching threshold asymmetric GARCH model with Studentt innovations. *The Econometrics Journal*, 12(1):105–126.

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.

Fruehwirth-Schnatter, S. and Frühwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics & Data Analysis*, 51(7):3509–3528.

Frühwirth-Schnatter, S. and Frühwirth, R. (2010). Data augmentation and mcmc for binary and multinomial logit models. In *Statistical modelling and regression structures*, pages 111–132. Springer.

Frühwirth-Schnatter, S., Frühwirth, R., Held, L., and Rue, H. (2009). Improved auxiliary mixture sampling for hierarchical models of non-gaussian data. *Statistics and Computing*, 19(4):479–492.

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, 93(4):827–841.

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

Fussl, A., Fruehwirth-Schnatter, S., and Fruehwirth, R. (2013). Efficient mcmc for binomial logit models. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 23(1):1–21.

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.

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

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

Gómez-Rubio, V. and Rue, H. (2018). Markov chain monte carlo with the integrated nested laplace approximation. *Statistics and Computing*, 28(5):1033–1051.

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.

Neal, P. and Subba Rao, T. (2007). Mcmc for integer-valued arma processes. *Journal of Time Series Analysis*, 28(1):92–110.

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.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.

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 districutions 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.

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

Weiß, C. H. (2009). Modelling time series of counts with overdispersion. *Statistical Methods 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.

Wong, T.-T. (1998). Generalized dirichlet distribution in bayesian analysis. *Applied Mathematics and Computation*, 97(2-3):165–181.

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.

2.8 Appendix

2.8.1 Distributions used in this chapter

Poisson Difference distribution

The Poisson difference distribution, a.k.a. 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

$$P(X = x) = e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{x/2} I_{|x|}(2\sqrt{\lambda_1\lambda_2}), \qquad \text{with } X \in \mathbb{Z} \tag{2.63}$$

where $\mathbb{Z} = \ldots, -1, 0, 1, \ldots$ 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)

$$I_v(z) = \left(\frac{z}{2}\right)^2 \sum_{k=0}^{\infty} \frac{\left(\frac{z^2}{4}\right)^k}{k!\Gamma(v+k+1)} \tag{2.64}$$

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: $\sum_{k=-\infty}^{+\infty} p(k; \lambda_1, \lambda_2) = 1$

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

$$P_x(\theta,\lambda) = \begin{cases} \frac{\theta(\theta+\lambda x)^{x-1}}{x!} e^{(-\theta-\lambda x)}, & x = 0,1,2,\ldots \\ 0, & \text{for } x > m \text{ if } \lambda < 0. \end{cases} \tag{2.65}$$

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:

$$z = \sum_{x=1}^{\infty} \frac{u^x}{x!} \{D^{x-1}(g(z))^x\}_{z=0} \tag{2.66}$$

$$\begin{aligned} f(z) &= \sum_{x=0}^{\infty} \frac{u^x}{x!} D^{x-1}\{(g(z))^x f'(z)\}|_{z=0} \\ &= f(0) + \sum_{x=1}^{\infty} \frac{u^x}{x!} D^{x-1}\{(g(z))^x f'(z)\}|_{z=0} \end{aligned} \tag{2.67}$$

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

$$f(z) = e^{\theta(z-1)} \qquad \text{and} \qquad g(z) = e^{\lambda(z-1)} \tag{2.68}$$

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:

$$(\theta - e^{-\lambda})(e^{\lambda} - 2\lambda)^{-1} < M < a \tag{2.69}$$

where a is the smallest value of M satisfying the inequality

$$\lambda^2 M^2 + M[2\lambda\theta - (\theta + 2\lambda)e^{\lambda}] > 0 \tag{2.70}$$

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:

$$(1 - \lambda)\mu'_{k+1} = \theta\mu'_k + \theta\frac{\partial\mu'_k}{\partial\theta} + \lambda\frac{\partial\mu'_k}{\partial\lambda}, \qquad k = 0, 1, 2, \ldots \tag{2.71}$$

$$(1 - \lambda)K_{k+1} = \lambda\frac{\partial K_k}{\partial\lambda} + \theta\frac{\partial K_k}{\partial\theta} +, \qquad k = 1, 2, 3, \ldots. \tag{2.72}$$

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

$$\mu_{k+1} = \frac{\theta k}{(1-\lambda)^3}\mu_{k-1} + \frac{1}{1-\lambda}\left\{\frac{d\,\mu_k(t)}{dt}\right\}_{t=1}, \qquad k = 1, 2, 3, \ldots \tag{2.73}$$

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

Generalized Poisson Difference distribution

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

$$P_x(\theta, \lambda) = \frac{\theta(\theta + x\lambda)^{x-1}}{x!}e^{-\theta - x\lambda} \qquad x = 0, 1, 2, \ldots \tag{2.74}$$

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:

$$P(D = d) = e^{-\theta_1 - \theta_2 - d\lambda}\sum_{y=0}^{\infty}\frac{\theta_2(\theta_2 + y\lambda)^{y-1}}{y!}\frac{\theta_1(\theta_1 + (y+d)\lambda)^{y+d-1}}{(y+d)!}e^{-2y\lambda} \tag{2.75}$$

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 11. *The probability distribution in (2.75) can be written as*

$$P(D=d) = e^{-(\theta_1+\theta_2)} \cdot e^{-d\lambda} \sum_{y=0}^{\infty} \frac{\theta_1\theta_2}{y!(y+d)!}(\theta_2+y\lambda)^{y-1}(\theta_1+(y+d)\lambda)^{y+d-1} \cdot e^{-2y\lambda} \tag{2.76}$$

$$= C_d(\theta_1, \theta_2, \lambda). \tag{2.77}$$

Therefore, equation (2.76) is the pgf of the difference of two GP variates, from which is possible to obtain the following particular cases:

$$\begin{cases} C_d(\theta_1, 0, \lambda) = \frac{\theta_1(\theta_1+d\lambda)^{d-1}}{d!}e^{-\theta_1-d\lambda}, & \text{for } d = 0, 1, 2, \dots \\ C_d(0, \theta_2, \lambda) = \frac{\theta_2(\theta_2-d\lambda)^{-d-1}}{(-d)!}e^{-\theta_2+d\lambda}, & \text{for } d = 0, -1, -2, \dots \\ C_d(\theta_1, \theta_2, 0) = e^{-\theta_1-\theta_2}(\theta_1)^{d/2}(\theta_2)^{-d/2}I_d(2\sqrt{\theta_1\theta_2}), & \end{cases} \tag{2.78}$$

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 (2.78) 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

$$G(u) = \exp[(\theta_1(t_1-1) + (\theta_2(t_2-1)]. \tag{2.79}$$

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:

$$L_1 = \frac{(\theta_1-\theta_2)}{1-\lambda} \text{ is the first cumulant and the mean.} \tag{2.80}$$

$$L_2 = \frac{(\theta_1+\theta_2)}{(1-\lambda)^3} \text{ is the second cumulant and the variance.} \tag{2.81}$$

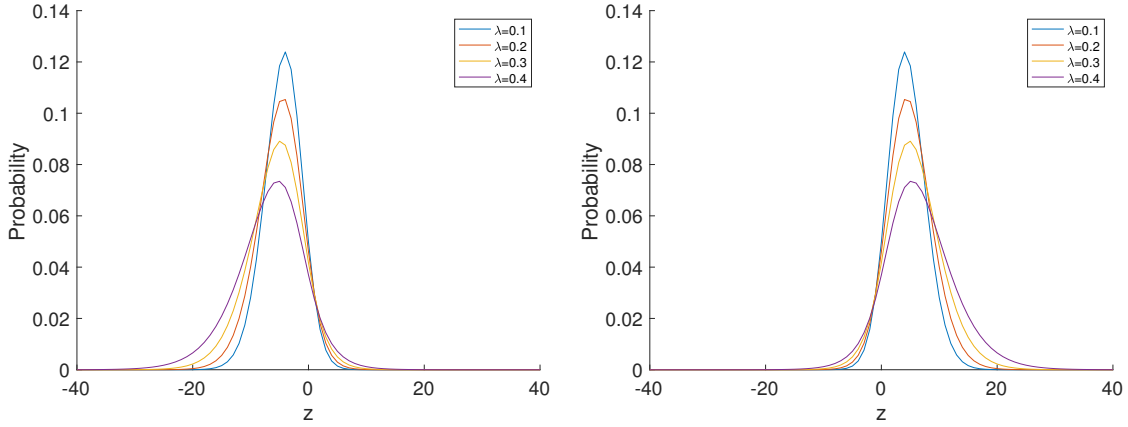
$$\beta_1 = \frac{(\theta_1 - \theta_2)^2}{(\theta_1 + \theta_2)^3} \frac{(1 + 2\lambda)^2}{1 - \lambda} \text{ is the skewness.} \tag{2.82}$$

$$\beta_2 = 3 + \frac{1 + 8\lambda + 6\lambda^2}{(\theta_1 + \theta_2)(1 - \lambda)} \text{ is the kurtosis.} \tag{2.83}$$

In Fig. 2.13-2.14, we show the GPD for various setting of λ, σ^2 and μ.

(a) GPD for $\mu = -4$ and $\sigma^2 = 8$

(b) GPD for $\mu = 4$ and $\sigma^2 = 8$



(c) GPD for $\mu = 2$ and $\sigma^2 = 5$

(d) GPD for $\mu = 2$ and $\sigma^2 = 15$

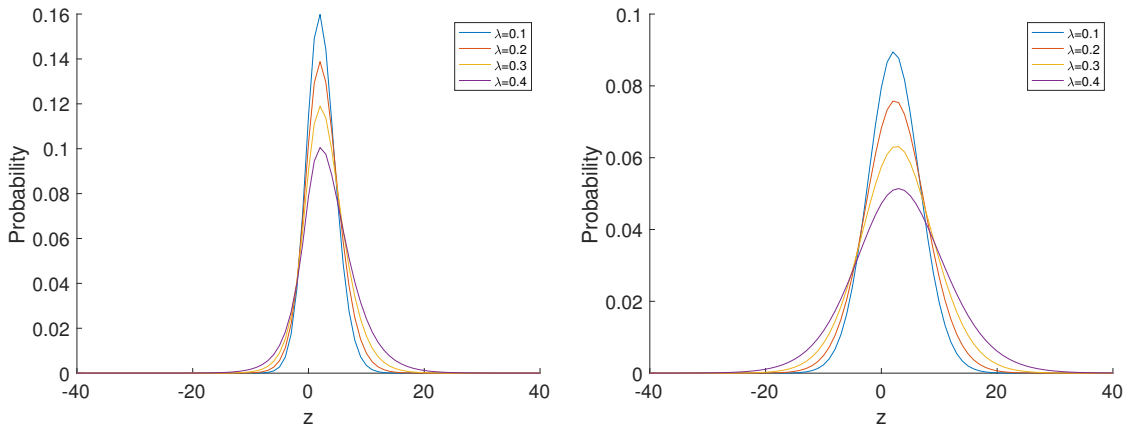


Figure 2.13: Generalized Poisson difference distribution for different values of λ. Panels (a) and (b) show the GPD when λ varies, for a fixed value of $\sigma^2 = 8$ and two different values of μ. Panels (c) and (d) show the GPD when λ varies, for a fixed value of $\mu = 2$ and two different value of σ^2.

(a) GPD for different values of μ (b) GPD for different values of σ^2

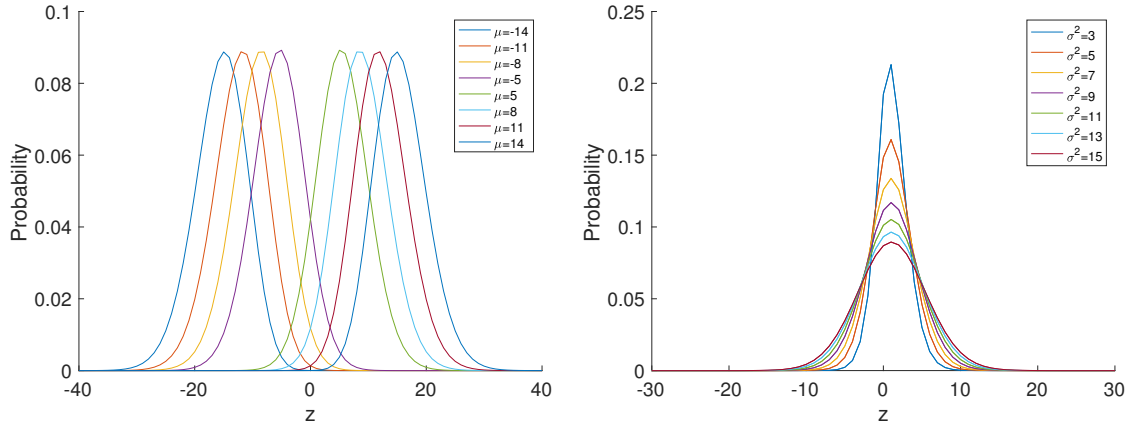


Figure 2.14: Generalized poisson difference distribution for different values of μ and σ and fixed λ. In panel (a) μ varies, while $\sigma^2 = 15$ and $\lambda = 0.1$ are fixed. In panel (b) σ^2 varies and $\mu = 1$ and $\lambda = 0.1$ are fixed.

Figure 2.13 shows how the GPD distribution varies when λ varies. We set $\lambda = (0, 0.1, 0.2, 0.3, 0.4)$ since smaller values and, possibly, negative values do not met 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 2.14 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.

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 byConsul and Famoye (2006) as QB-II.

$$P(X = x) = \binom{n}{x} \frac{ab}{a+b} \frac{(a + x\theta)^{x-1}(b + n\theta - x\theta)^{n-x-1}}{(a + b + n\theta)^{n-1}} \tag{2.84}$$

for $x = 0, 1, 2, \ldots, n$ 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:

$$P(X = x) = pq\binom{n}{x}\frac{(p + x\theta)^{x-1}(q + (n - x)\theta)^{n-x-1}}{(1 + n\theta)^{n-1}} \tag{2.85}$$

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 QB(p,θ,n). For large n, such that $np \to \lambda$, 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 12 (**Quasi-Binomial Thinning**). *Let $\rho_{\theta,\lambda}\circ$ be the quasi-binomial thinning operator such that $\rho_{\theta,\lambda}\circ$ follows a QB(ρ,θ/λ,x). 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($\rho\lambda$,θ) 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: $E[\rho_{\theta,\lambda} \circ X, X] = \rho \cdot \sigma_X^2$

2.8.2 Proofs of the results of the chapter

Proofs of the results in Section 2.2

Proof of Lemma 1. 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

$$X = X_1 - X_2 \text{ and } Y = Y_1 - Y_2 \tag{2.86}$$

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

$$\begin{aligned} X + Y &= X_1 - X_2 + Y_1 - Y_2 \\ &= (X_1 + Y_1) - (X_2 + Y_2) \\ &= GP(\theta_1 + \theta_3, \lambda) - GP(\theta_2 + \theta_4, \lambda) \sim GPD(\theta_1 + \theta_3, \theta_2 + \theta_4, \lambda). \end{aligned} \tag{2.87}$$

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

$$\sum_{i=1}^{n} Z_i \sim GPD(\sum_{i=1}^{n} \theta_{1i}, \sum_{i=1}^{n} \theta_{2i}, \lambda). \tag{2.88}$$

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

$$X = X_1 - X_2 \text{ and } Y = Y_1 - Y_2 \tag{2.89}$$

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

$$\begin{aligned} X - Y &= X_1 - X_2 - Y_1 + Y_2 \\ &= (X_1 + Y_2) - (X_2 + Y_1) \\ &= GP(\theta_1 + \theta_4, \lambda) - GP(\theta_2 + \theta_3, \lambda) \sim GPD(\theta_1 + \theta_4, \theta_2 + \theta_3, \lambda). \end{aligned} \tag{2.90}$$

□

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:

$$P(Z = z) = e^{-(\theta_1+\theta_2)} \cdot e^{-z\lambda} \sum_{s=max(0,-z)}^{+\infty} \frac{\theta_1\theta_2}{s!(s+z)!} (\theta_2+\lambda s)^{s-1} (\theta_1+\lambda(s+z))^{s+z-1} \cdot e^{-2\lambda s} \tag{2.91}$$

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

$$\begin{cases} \mu = \theta_1 - \theta_2 \\ \sigma^2 = \theta_1 + \theta_2 \end{cases}, \tag{2.92}$$

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

$$\begin{cases} \theta_1 = \frac{\sigma^2+\mu}{2} \\ \theta_2 = \frac{\sigma^2-\mu}{2} \end{cases} \tag{2.93}$$

By substituting 2.93 into equation 2.91 we have

$$\begin{aligned} P(Z = z) \quad = \quad & e^{(\frac{\sigma^2+\mu}{2} + \frac{\sigma^2-\mu}{2})} e^{-z\lambda} \sum_{s=max(0,-z)}^{+\infty} \frac{\left(\frac{\sigma^2+\mu}{2}\right)\left(\frac{\sigma^2-\mu}{2}\right)}{s!(s+z)!} \left[\frac{\sigma^2+\mu}{2} + (s+z)\lambda\right]^{s+z-1} \\ & \left[\frac{\sigma^2-\mu}{2} + s\lambda\right]^{s-1} e^{-2\lambda s} \end{aligned} \tag{2.94}$$

Carrying out the operations in Eq. 2.94 we obtain Eq. 2.92. □

Proof of Remark 2. If $Z \sim GPD(\theta_1, \theta_2, \lambda)$ in the parametrization of Consul (1986), the moments are given in Appendix 2.8.1. By using our reparametrization of the GPD

$$\theta_1 = \frac{\sigma^2 + \mu}{2}, \quad \theta_2 = \frac{\sigma^2 - \mu}{2} \tag{2.95}$$

we obtain mean, variance, skewness and kurtosis:

$$\begin{aligned} E(Z) &= \frac{\theta_1 - \theta_2}{1-\lambda} = \frac{\mu}{1-\lambda} \\ V(Z) &= \frac{\theta_1 + \theta_2}{(1-\lambda)^3} = \frac{\sigma^2}{(1-\lambda)^3} = \kappa_2 \\ S(Z) &= \frac{\mu^{(3)}}{\sigma^3} = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{\mu}{\sigma^3}\frac{(1+2\lambda)}{\sqrt{1-\lambda}} \\ K(Z) &= \frac{\kappa_4 + 3\kappa_2^2}{\sigma^4} = 3 + \frac{1 + 8\lambda + 6\lambda^2}{\sigma^2(1-\lambda)} \end{aligned} \tag{2.96}$$

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

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

$$\begin{aligned} \mu &= E[Z_t] = E[E[Z_t|\mathcal{F}_{t-1}]] \\ &= E\left[\alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j Z_{t-j}\right] = \alpha_0 D^{-1}(1) + \mu D^{-1}(1) G(1) \end{aligned} \tag{2.97}$$

$$\Leftrightarrow \mu = \frac{\alpha_0}{D(1) - G(1)} = \alpha_0 \left(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j\right)^{-1} = \alpha_0 K^{-1}(1). \tag{2.98}$$

Where we go from line two to line three of eq 2.97 as follows

$$\begin{aligned} E[Z_t|\mathcal{F}_{t-1}] &= E[\alpha_0 D^{-1}(1) + H(B) Z_t|\mathcal{F}_{t-1}] \\ &= E[\alpha_0 D^{-1}(1)] + E[H(B) Z_t|\mathcal{F}_{t-1}] \\ &= \alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j Z_{t-j}. \end{aligned} \tag{2.99}$$

Following Ferland et al. (2006), to go from line three to line four of 2.97:

$$\begin{aligned} E[\alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j Z_{t-j}] &= E[\alpha_0 D^{-1}(1)] + E[\sum_{j=1}^{\infty} \psi_j Z_{t-j}] \\ &= \alpha_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j E[Z_{t-j}] \\ &= \alpha_0 D^{-1}(1) + \mu H(1) \\ &= \alpha_0 D^{-1}(1) + \mu D^{-1}(1) G(1). \end{aligned} \tag{2.100}$$

From 2.98, a necessary condition for the second-order stationarity of the integer-valued process $\{Z_t\}$ is: $\left(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j\right) > 0$. □

Proof of the results in Section 2.3

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. From Zhu (2012) and Ferland et al. (2006), we know that a sequence of GP random processes $X_t^{(n)}$ has an almost sure limit and that is X_t. 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

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

and

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

Let

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

and

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

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

$$
\begin{aligned}
\int_{\Omega} \mathbb{I}_B(\omega) d\mathbb{P}(\omega) &= \int_{(\Omega\cap A)\cup(\Omega\cap A^C)} \mathbb{I}_B(\omega) d\mathbb{P}(\omega) \\
&= \int_{\Omega} \mathbb{I}_B(\omega)\mathbb{I}_A(\omega) d\mathbb{P}(\omega) + \int_{\Omega} \mathbb{I}_B(\omega)\mathbb{I}_{A^C}(\omega) d\mathbb{P}(\omega) \\
&= \int_{\Omega} \mathbb{I}_B(\omega)\mathbb{I}_A(\omega) d\mathbb{P}(\omega) + \int_{\Omega} \mathbb{I}_{A^C}(\omega) \left(\int_{\Omega} \mathbb{I}_{B\cap A^C}(\omega) d\mathbb{P}(\omega|\omega') \right) d\mathbb{P}(\omega') \\
&= \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^C) \underbrace{\mathbb{P}(A^C)}_{=0} \\
&= \mathbb{P}(\{\omega : \lim_{n\to\infty} (aX_n(\omega) + bY_n(\omega)) = aX(\omega) + bY(\omega)\}|A) \underbrace{\mathbb{P}(A)}_{=1} \\
&= \mathbb{P}(\{\omega : a \lim_{n\to\infty} X_n(\omega) = aX(\omega) - bY(\omega) + bY(\omega)\}) = 1
\end{aligned}
\tag{2.101}
$$

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 $Z_t^{(n)} = X_t^{(n)} - Y_t^{(n)}$. □

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*

$$
\begin{aligned}
X_n(\omega) &\xrightarrow{L^2} X(\omega) \\
Y_n(\omega) &\xrightarrow{L^2} Y(\omega)
\end{aligned}
\tag{2.102}
$$

also their sum will have a mean-square limit.

$$
aX_n(\omega) + bY_n(\omega) \xrightarrow{L^2} aX(\omega) + bY(\omega), \quad \forall a, b \in \mathbb{R} \tag{2.103}
$$

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

$$
X_n(\omega) - Y_n(\omega) \xrightarrow{L^2} X(\omega) - Y(\omega) \tag{2.104}
$$

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

Proof of Proposition 6.

$$
g_{\mathbf{Z}}(\mathbf{t}) = \mathbb{E}\left[\prod_{i=1}^{k} t_i^{Z_i}\right] = \mathbb{E}\left[\prod_{i=1}^{k} t_i^{X_i}\right] \mathbb{E}\left[\prod_{i=1}^{k} \frac{1}{t_i^{Y_i}}\right] = g_{\mathbf{X}}(\mathbf{t}) g_{\mathbf{Y}}(\mathbf{t^{-1}}) \tag{2.105}
$$

□

Proof of Proposition 7. 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

$$\mathbf{Z}^{(n)}_{1+h\ldots k+h} = (Z^{(n)}_{1+h}, \ldots, Z^{(n)}_{k+h})' \text{ and } \mathbf{Z}^{(n)}_{1\ldots k} = (Z^{(n)}_{1}, \ldots, Z^{(n)}_{k})' \tag{2.106}$$

have the same joint distribution, where we can rewrite both vectors in Eq. 2.106 as

$$\begin{aligned} \mathbf{Z}^{(n)}_{1+h\ldots k+h} &= (\mathbf{X}^{(n)}_{1+h\ldots k+h} - \mathbf{Y}^{(n)}_{1+h\ldots k+h})' \\ &= ((X^{(n)}_{1+h} - Y^{(n)}_{1+h}), \ldots, (X^{(n)}_{k+h} - Y^{(n)}_{k+h}))' \end{aligned} \tag{2.107}$$

and

$$\begin{aligned} \mathbf{Z}^{(n)}_{1\ldots k} &= (\mathbf{X}^{(n)}_{1\ldots k} - \mathbf{Y}^{(n)}_{1\ldots k})' \\ &= ((X^{(n)}_{1} - Y^{(n)}_{1}), \ldots, (X^{(n)}_{k} - Y^{(n)}_{k}))' \end{aligned} \tag{2.108}$$

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.

$$\begin{aligned} g_{\mathbf{X}^{(n)}_{1\ldots k}}(\mathbf{t}) &= E\left[\prod_{j=1}^{k} t_j^{X_j^{(n)}}\right] = E\left[E_{\mathbf{X}^{(n)}_{1\ldots k}|\mathbf{U}_{1,1-n\ldots k}}\left[\prod_{j=1}^{k} t_j^{X_j^{(n)}}\right]\right] \\ &= \sum_{\mathbf{v_1}\in\mathbb{N}^{(k+n)}} E_{\mathbf{X}^{(n)}_{1\ldots k}|\mathbf{U}_{1,1-n\ldots k}=\mathbf{v}_1}\left[\prod_{j=1}^{k} t_j^{X_j^{(n)}}\right] Pr\left(\mathbf{U}_{1,1-n\ldots k} = \mathbf{v}_1\right) \end{aligned} \tag{2.109}$$

$$g_{\mathbf{Y}^{(n)}_{1\ldots k}}(\mathbf{t}) = \sum_{\mathbf{v_2}\in\mathbb{N}^{(k+n)}} E_{\mathbf{Y}^{(n)}_{1\ldots k}|\mathbf{U}_{2,1-n\ldots k}=\mathbf{v}_2}\left[\prod_{j=1}^{k} t_j^{Y_j^{(n)}}\right] Pr\left(\mathbf{U}_{2,1-n\ldots k} = \mathbf{v}_2\right) \tag{2.110}$$

$$G_{\mathbf{Z}^{(n)}_{1\ldots k}}(\mathbf{t}) = \sum_{\mathbf{v}\in\mathbb{N}^{(k+n)}} E_{\mathbf{Z}^{(n)}_{1\ldots k}|\mathbf{U}_{1-n\ldots k}=\mathbf{v}}\left[\prod_{j=1}^{k} t_j^{(X_j^{(n)}-Y_j^{(n)})}\right] Pr\left(\mathbf{U}_{1-n\ldots k} = \mathbf{v}\right) \tag{2.111}$$

By the thinning representation, for a any given value $\mathbf{u}_{1,t-n\ldots t+k} = (u_{1,t-n}, \ldots, u_{1,t+k})'$ of the vector $\mathbf{U}_{1,t-n\ldots t+k} = (U_{1,t-n}, \ldots, U_{1,t+k})'$ and $\mathbf{u}_{2,t-n\ldots t+k} = (u_{2,t-n}, \ldots, u_{2,t+k})'$ of the vector $\mathbf{U}_{2,t-n\ldots t+k} = (U_{2,t-n}, \ldots, U_{2,t+k})'$, the components of the vectors $(X_1^{(n)}, \ldots, X_k^{(n)})'$ and $(Y_1^{(n)}, \ldots, Y_k^{(n)})'$ are computed using a set of well-determined variables from the sequences $\mathcal{V}_{1,\tau,\eta}$ and $\mathcal{V}_{2,\tau,\eta}$, where $\tau = t-n, \ldots, t+k-1$ and $\eta = 1, \ldots, n$. Therefore, if $\mathbf{U}_{1,t-n\ldots t+k}$ and $\mathbf{U}_{1,t-n+h\ldots t+k+h}$ are both fixed to the same value $\mathbf{v_1}$ and $\mathbf{U}_{2,t-n\ldots t+k}$ and $\mathbf{U}_{2,t-n+h\ldots t+k+h}$ are both fixed to the same value $\mathbf{v_2}$, it follows that the conditional distribution of

$$\mathbf{Z}^{(n)}_{1+h\ldots k+h} = ((X^{(n)}_{1+h} - Y^{(n)}_{1+h}), \ldots, (X^{(n)}_{k+h} - Y^{(n)}_{k+h}))'$$

and

$$\mathbf{Z}^{(n)}_{1\ldots k} = ((X^{(n)}_{1} - Y^{(n)}_{1}), \ldots, (X^{(n)}_{k} - Y^{(n)}_{k}))'$$

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

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

and, since

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

it is possible to write

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

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

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 Generelized 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,

$$E[Z_t] = E[X_t] - E[Y_t] = \mu_X - \mu_Y \leq \mu_X + \mu_Y \leq C_1 + C_1' \tag{2.112}$$

is finite and

$$V[X_t - Y_t] = V[X_t] + V[X_t] = \sigma_X^2 + \sigma_Y^2 \leq C_2 + C_2' \tag{2.113}$$

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 Proposition 10. 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

$$\mu_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i} + \sum_{j=1}^{q} b_j \mu_{t-j}. \tag{2.114}$$

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

$$r_{1t}^{(n)} = (1-\lambda)U_{1t} + (1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{X_{t-i}} V_{1t-i,i,k} \tag{2.115}$$

$$r_{2t}^{(n)} = (1-\lambda)U_{2t} + (1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{Y_{t-i}} V_{2t-i,i,k}. \tag{2.116}$$

and

$$r_t^{(n)} = r_{1t}^{(n)} - r_{2t}^{(n)}. \tag{2.117}$$

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

$$X_t - r_{1t}^{(n)} = (X_t - X_t^{(n)}) + (X_t^{(n)} - r_{1t}^{(n)}) \tag{2.118}$$

and

$$Y_t - r_{2t}^{(n)} = (Y_t - Y_t^{(n)}) + (Y_t^{(n)} - r_{2t}^{(n)}). \tag{2.119}$$

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. 2.118 and 2.119 goes to zero. Therefore, we can write

$$\begin{aligned} Z_t - r_t^{(n)} &= (X_t - Y_t) - (r_{1t}^{(n)} - r_{2t}^{(n)}) \\ &= \left[(X_t - X_t^{(n)}) - (Y_t - Y_t^{(n)})\right] + \left[(X_t^{(n)} - r_{1t}^{(n)}) - (Y_t^{(n)} - r_{2t}^{(n)})\right] \\ &= (Z_t - Z_t^{(n)}) + \left[(X_t^{(n)} - Y_t^{(n)}) - (r_{1t}^{(n)} - r_{2t}^{(n)})\right] \\ &= (Z_t - Z_t^{(n)}) + \left[Z_t^{(n)} - (r_{1t}^{(n)} - r_{2t}^{(n)})\right], \end{aligned} \tag{2.120}$$

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. 2.120 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

$$\begin{aligned} W_t^{(n)} &= (r_{1t}^{(n)} - r_{2t}^{(n)}) - (X_t - Y_t) \\ &= (r_{1t}^{(n)} - X_t) - (r_{2t}^{(n)} - Y_t) \end{aligned} \tag{2.121}$$

Ferland et al. (2006) show that

$$\begin{aligned} \lim_{n\to\infty} E\left[(r_{1t}^{(n)} - X_t)\right] = 0 \\ \lim_{n\to\infty} E\left[(r_{2t}^{(n)} - Y_t)\right] = 0 \end{aligned} \tag{2.122}$$

therefore, we can conclude that also

$$\lim_{n\to\infty} E\left[(r_t^{(n)} - Z_t)\right] = 0. \tag{2.123}$$

Equation 2.123 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 that $r_t^{(n_k)} \xrightarrow{a.s.} Z_t$. 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

$$r_t^{(n)}|\mathcal{F}_{t-1} \sim GPD\left(\alpha_0 D^{-1}(1) + \sum_{j=1}^{n} \psi_j Z_{t-j}\right) \tag{2.124}$$

and conclude that

$$Z_t|\mathcal{F}_{t-1} \sim GPD(\tilde{\mu}_t, \tilde{\sigma}_t^2, \lambda). \tag{2.125}$$

□

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

$$\gamma_Z(k) = \alpha_1 \gamma_Z(k-1) = \alpha_1^{k-1}\gamma_Z(1) \tag{2.126}$$

For $k = 1$:

$$\gamma_Z(1) = Cov(Z_t, Z_{t-1} = \alpha_1\gamma_Z(0) = \alpha_1 V(Z_t = \alpha_1[\phi^3 E(\sigma_t^{2^*})] + \alpha_1 V(\mu_t). \tag{2.127}$$

For $k \geq 1$ we have

$$\gamma_\mu(k) = \alpha_1\gamma_\mu(k-1) = \alpha_1^k V(\mu_t). \tag{2.128}$$

For $k = 0$:

$$\begin{aligned} \gamma_\mu(k) = V(\mu_t) &= \alpha_1\gamma_Z(1) \\ &= \alpha_1\left\{\alpha_1[\phi^3 E(\sigma_t^{2^*}) + V(\mu_t)]\right\} = \alpha_1^2[\phi^3 E(\sigma_t^{2^*})] + \alpha_1^2 V(\mu_t) \end{aligned} \tag{2.129}$$

Therefore,

$$V(\mu_t) = \frac{\alpha_1^2[\phi^3 E(\sigma_t^{2^*})]}{1-\alpha_1^2} \tag{2.130}$$

and

$$\begin{aligned} V(Z_t) &= \phi^3 E(\sigma_t^{2^*}) + V(\mu_t) \\ &= \phi^3 E(\sigma_t^{2^*}) + \frac{\alpha_1^2[\phi^3 E(\sigma_t^{2^*})]}{1-\alpha_1^2} = \frac{\phi^3 E(\sigma_t^{2^*})}{1-\alpha_1^2} \end{aligned} \tag{2.131}$$

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

$$\rho_\mu(k) = \frac{\gamma_\mu(k)}{V(\mu_t)} = \frac{\alpha_1^k V(\mu_t)}{V(\mu_t)} = \alpha_1^k \tag{2.132}$$

$$\begin{aligned}\rho_Z(k) &= \frac{\gamma_Z(k)}{V(Z_t)} = \alpha_1^{k-1}\gamma_Z(1)\frac{1-\alpha_1^2}{\phi^3 E(\sigma_t^{2^*})} \\ &= \alpha_1^{k-1}\frac{\alpha_1(1-\alpha_1^2)\phi^3 E(\sigma_t^{2^*}) + \alpha_1^3\phi^3 E(\sigma_t^{2^*})}{1-\alpha_1^2}\frac{1-\alpha_1^2}{\phi^3 E(\sigma_t^{2^*})} = \alpha_1^k\end{aligned} \tag{2.133}$$

□

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

$$\gamma_Z(k) = \alpha_1\gamma_Z(k-1) + \beta_1\gamma_Z(k-1) = (\alpha_1+\beta_1)^{k-1}\gamma_Z(1) \tag{2.134}$$

For $k = 1$:

$$\begin{aligned}\gamma_Z(1) &= Cov(Z_t, Z_{t-1}) = \alpha_1\gamma_Z(0) + \beta_1\gamma_\mu(0) \\ &= \alpha_1 V(Z_t) + \beta_1 V(\mu_t) = \alpha_1[\phi^3 E(\sigma_t^{2^*})] + (\alpha_1+\beta_1)V(\mu_t).\end{aligned} \tag{2.135}$$

For $k \geq 1$ we have

$$\gamma_\mu(k) = \alpha_1\gamma_\mu(k-1) + \beta_1\gamma_\mu(k-1) = (\alpha_1+\beta_1)^k V(\mu_t). \tag{2.136}$$

For $k = 0$:

$$\begin{aligned}\gamma_\mu(k) &= V(\mu_t) = \alpha_1\gamma_Z(1) + \beta_1\gamma_\mu(1) \\ &= \alpha_1\left\{\alpha_1[\phi^3 E(\sigma_t^{2^*})] + (\alpha_1+\beta_1)V(\mu_t)\right\} + \beta_1\left[(\alpha_1+\beta_1)V(\mu_t)\right] \\ &= \alpha_1^2[\phi^3 E(\sigma_t^{2^*})] + (\alpha_1+\beta_1)^2 V(\mu_t)\end{aligned} \tag{2.137}$$

Therefore,

$$V(\mu_t) = \frac{\alpha_1^2[\phi^3 E(\sigma_t^{2^*})]}{1-(\alpha_1+\beta_1)^2} \tag{2.138}$$

and

$$\begin{aligned}V(Z_t) &= \phi^3 E(\sigma_t^{2^*}) + V(\mu_t) = \phi^3 E(\sigma_t^{2^*}) + \frac{\alpha_1^2[\phi^3 E(\sigma_t^{2^*})]}{1-(\alpha_1+\beta_1)^2} \\ &= \frac{\phi^3 E(\sigma_t^{2^*})[1-(\alpha_1+\beta_1)^2+\alpha_1^2]}{1-(\alpha_1+\beta_1)^2}\end{aligned} \tag{2.139}$$

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

$$\rho_\mu(k) = \frac{\gamma_\mu(k)}{V(\mu_t)} = \frac{(\alpha_1+\beta_1)^k V(\mu_t)}{V(\mu_t)} = (\alpha_1+\beta_1)^k \tag{2.140}$$

$$\begin{aligned}\rho_Z(k) &= \frac{\gamma_Z(k)}{V(Z_t)} = (\alpha_1+\beta_1)^{k-1}\gamma_Z(1)\frac{1-(\alpha_1+\beta_1)^2}{\phi^3 E(\sigma_t^{2^*})[1-(\alpha_1+\beta_1)^2+\alpha_1^2]} \\ &= (\alpha_1+\beta_1)^{k-1}\frac{\alpha_1[1-\beta_1(\alpha_1+\beta_1)]}{1-(\alpha_1+\beta_1)^2+\alpha_1^2}\end{aligned} \tag{2.141}$$

□

2.8.3 Further details for the numerical illustration

Correct implementation test

When programming a posterior simulator errors may arise, therefore we are interested in testing if the algorithm is mathematically correct or the mixing of the Markov chain. One technique to detect errors in the MCMC algorithm is proposed by Geweke (2004) and is a joint distribution test. Consider a model over parameters $\boldsymbol{\theta}$ and data $\boldsymbol{x}$, the idea is to test an MCMC sampler for the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$ by means of the joint density of the of the data and the parameters, that is $p(\boldsymbol{\theta}, \boldsymbol{y})$.

Let g be any function $g : \Theta \times Y$ such that

$$\int_{\Theta}\int_{Y} g^2(\boldsymbol{\theta}, \boldsymbol{y}) p(\boldsymbol{\theta}, \boldsymbol{y}) d\mu(\boldsymbol{y}) d\nu(\boldsymbol{\theta}) < \infty$$

that is $\sigma_g^2 = Var[g(\boldsymbol{\theta}, \boldsymbol{y})] < \infty$. Therefore, the joint distribution test compares two different simulation approximation of

$$\bar{g} = E[g(\boldsymbol{\theta}, \boldsymbol{y})] = \int_{\Theta}\int_{Y} g(\boldsymbol{\theta}, \boldsymbol{y}) p(\boldsymbol{\theta}, \boldsymbol{y}) d\mu(\boldsymbol{y}) d\nu(\boldsymbol{\theta}).$$

The two approximation use two simulator: (a) the marginal-conditional simulator and (b) the successive-conditional simulator.

The marginal-conditional simulator works as follows

$$\begin{aligned}
\boldsymbol{\theta}^{(m)} &\sim p(\boldsymbol{\theta}) \\
\boldsymbol{y}^{(m)} &\sim p(\boldsymbol{y}|\boldsymbol{\theta}^{(m)}) \\
g^{(m)} &= g(\boldsymbol{\theta}^{(m)}, \boldsymbol{y}^{(m)})
\end{aligned}$$

where $\bar{g}^M \xrightarrow{a.s.} \bar{g}$, $M^{1/2}(\bar{g}^M - \bar{g}) \xrightarrow{d} N(0, \sigma_g^2)$ and $\hat{\sigma}_g^{2(M)} \xrightarrow{a.s.} \sigma_g^2$.

For the successive-conditional simulator, first we need to draw $\boldsymbol{\theta}^{\widetilde{(0)}} \sim p(\boldsymbol{\theta})$, and then iterate the following steps

$$\begin{aligned}
\tilde{\boldsymbol{y}}^{(m)} &\sim p(\boldsymbol{y}|\tilde{\boldsymbol{\theta}}^{(m-1)}) \\
\tilde{\boldsymbol{\theta}}^{(m)} &\sim q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}^{(m-1)}, \tilde{\boldsymbol{y}}^{(m)}) \\
\tilde{g}^{(m)} &= g(\tilde{\boldsymbol{\theta}}^{(m)}, \tilde{\boldsymbol{y}}^{(m)})
\end{aligned}$$

where $\bar{\tilde{g}}^{(M)} \xrightarrow{a.s.} \bar{g}$ and $M^{1/2}(\bar{\tilde{g}}^{(M)}) - \bar{g} \xrightarrow{d} N(0, \tau_g^2)$.

Then, for $M_1 \to \infty$ and $M_2 \to \infty$, if the MCMC algorithm is error free, the statistics for the joint distribution test will be

$$Z = \frac{\bar{g}^{(M_1)} - \bar{\tilde{g}}^{(M_2)}}{(M_1^{-1}\hat{\sigma}_g^{2(M_1)} + M_2^{-1}\hat{\tau}_g^{2(M_1)})^{1/2}} \xrightarrow{d} N(0, 1). \tag{2.142}$$

Figure 2.15 shows the Geweke's statistics evaluated on the first n MCMC samples, $n = 1, \ldots, 2000$, for the parameters α, β and λ and the first three moments. A graphical inspection indicates convergence of the statistics. Table 2.5 reports the statistics on the 2000 samples for the different parameters (columns) and choices of the test function (rows). The absolute value of the statistics Z is always below 2.58 that is the critical value of the Geweke's statistics at the 1% level. Thus, the null hypothesis of correct implementation of the MCMC is accepted.

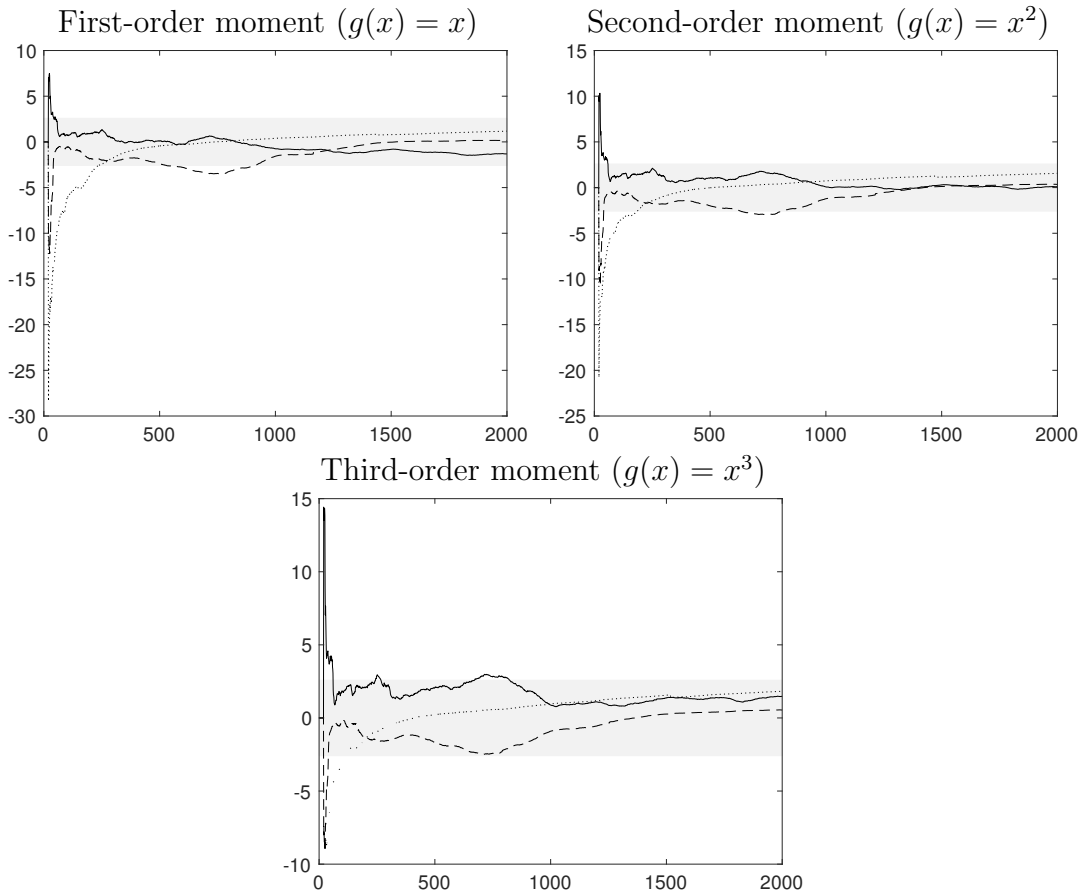


Figure 2.15: Plot of the Geweke's statistics for the correctness of the MCMC. The statistics is evaluated on the first n MCMC samples, $n = 1, \ldots, 2000$, for the parameters α (solid line), β (dashed line) and λ (dotted line) and the first three moments (different plots). Grey is the acceptance region of the test at the 1% level.

Sampling efficiency

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 $\lambda = 0.4$, $\alpha_1 = 0.25$, $\beta_1 = 0.23$, $\alpha_0 = -0.2$

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

Table 2.5: Value of the Geweke's test statistics for the correctness of the MCMC, for the three parameters of the GPD-INGARCH(1,1) for different choice of the test function $g(x)$.

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. 2.16-2.18 the MCMC output for one MCMC draw before removing the burn-in sample and thinning, while in Fig.2.19-2.21 the MCMC output after removing the burn-in sample and thinning.

2.8.4 Further details for the real-data applications

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

Table 2.6: Descriptive statistics, stationary tests and normality test for the accident and cyber datasets. ADF is the Augmented Dickey-Fuller test with stationarity as null hypothesis, PP is the Phillips-Perron test with stationarity as null hypothesis, KS is Kolmogorov-Smirnov test where the null hypothesis states that the data comes from a Normal distribution. The symbol ** means that the null hypothesis is rejected at 5% significance level.

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

Table 2.7: Descriptive statistics, stationary tests and normality test for the accident and cyber datasets. ADF is the Augmented Dickey-Fuller test with stationarity as null hypothesis, PP is the Phillips-Perron test with stationarity as null hypothesis, KS is Kolmogorov-Smirnov test where the null hypothesis states that the data comes from a Normal distribution. The symbol ** means that the null hypothesis is rejected at 5% significance level.

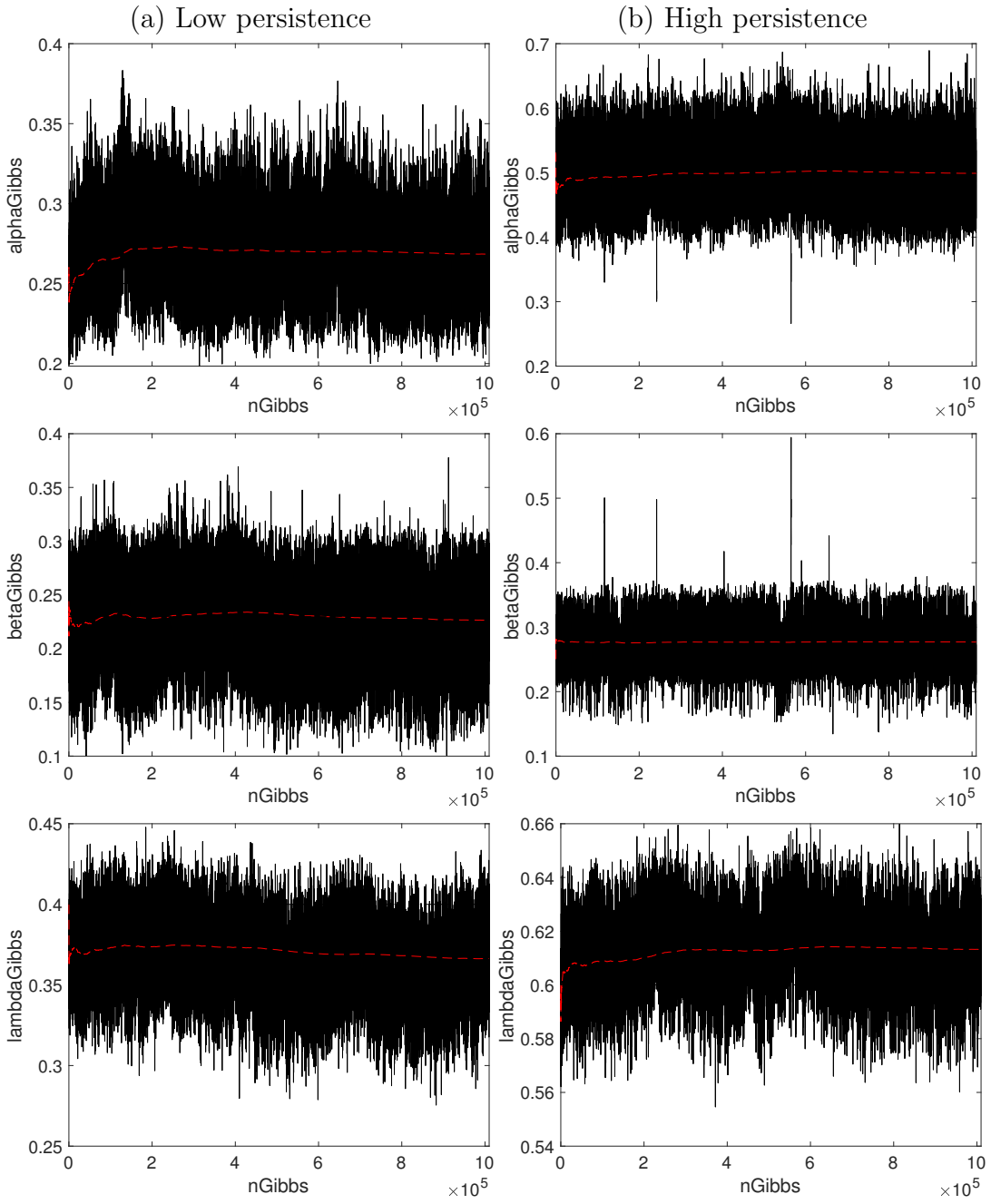


Figure 2.16: MCMC plot for the parameters in the two setting: low persistence and high persistence.

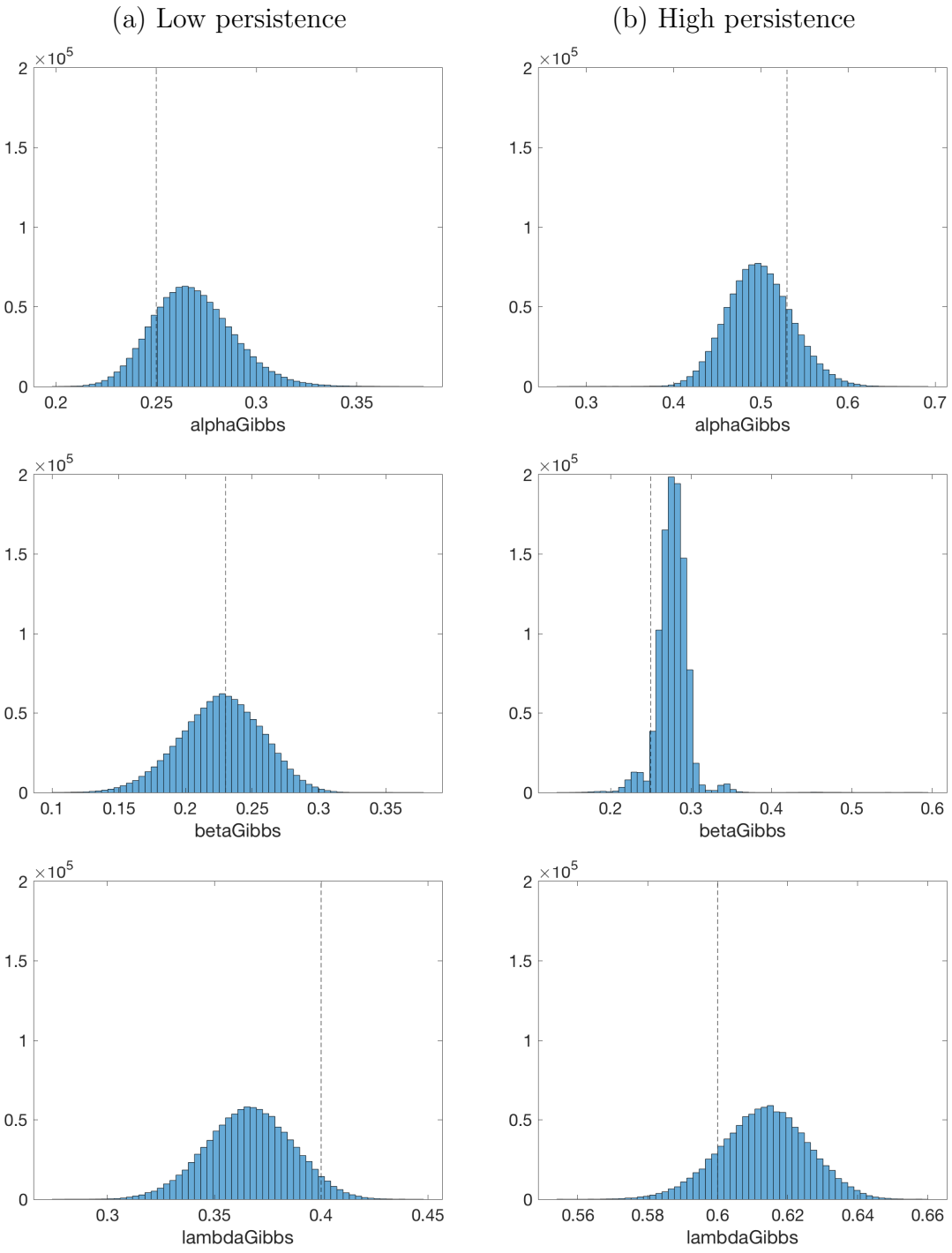


Figure 2.17: Histograms of the MCMC draws for the parameters in both settings: low persistence and high persistence.

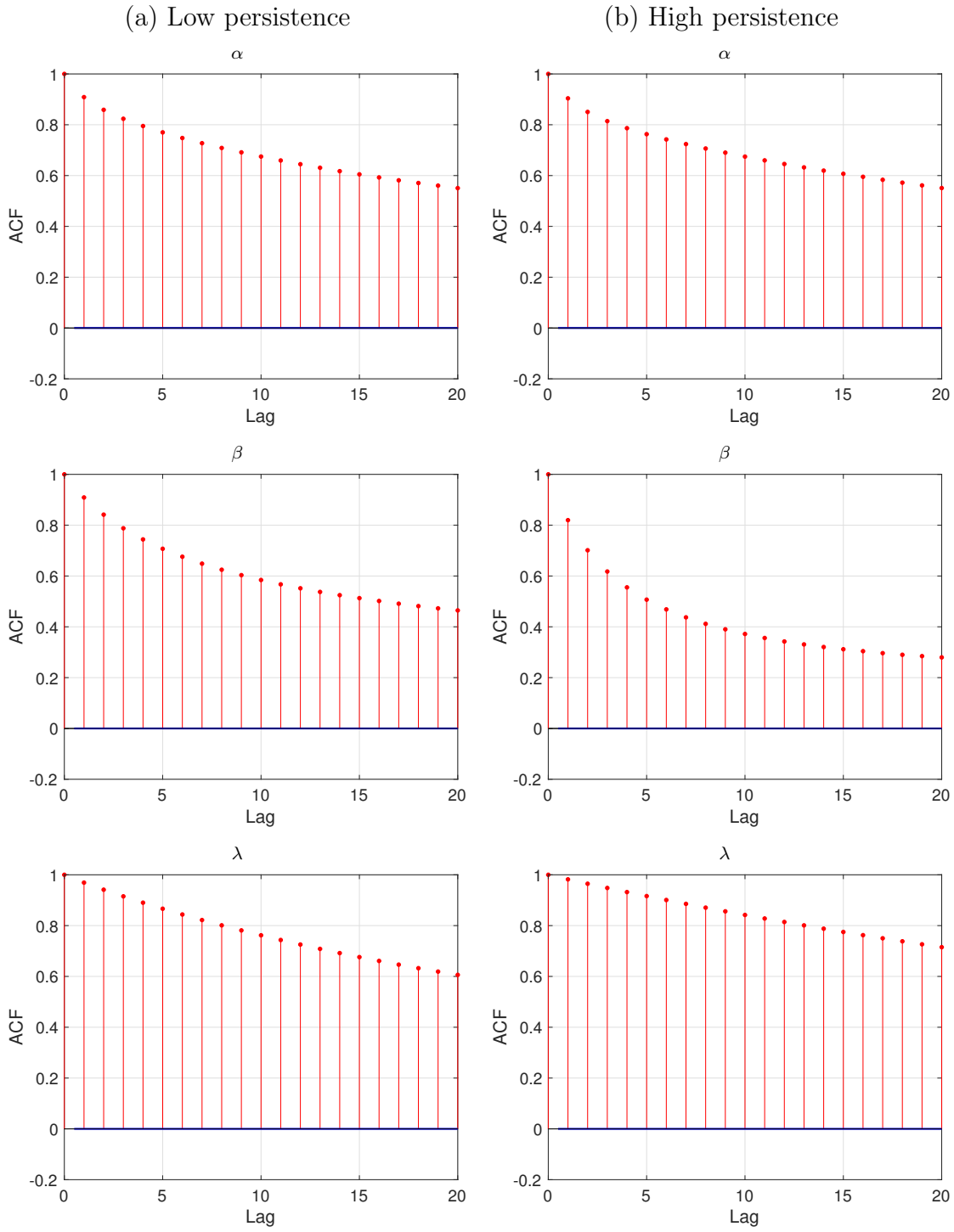


Figure 2.18: Autocorrelation function for the parameters in both low persistence and high persistence settings.

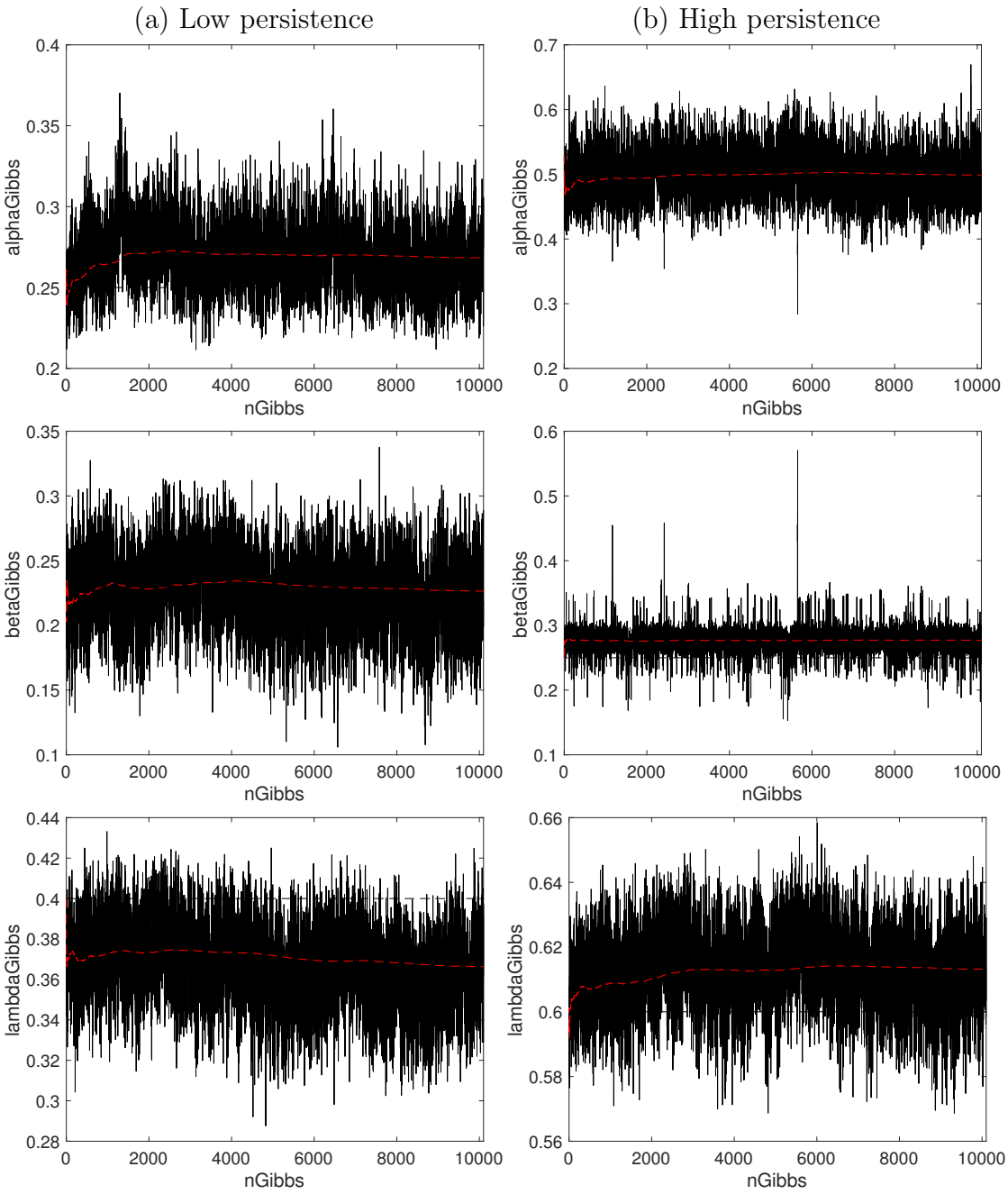


Figure 2.19: MCMC plot for the parameters in the two setting: low persistence and high persistence, after removing the burn-in sample and thinning.

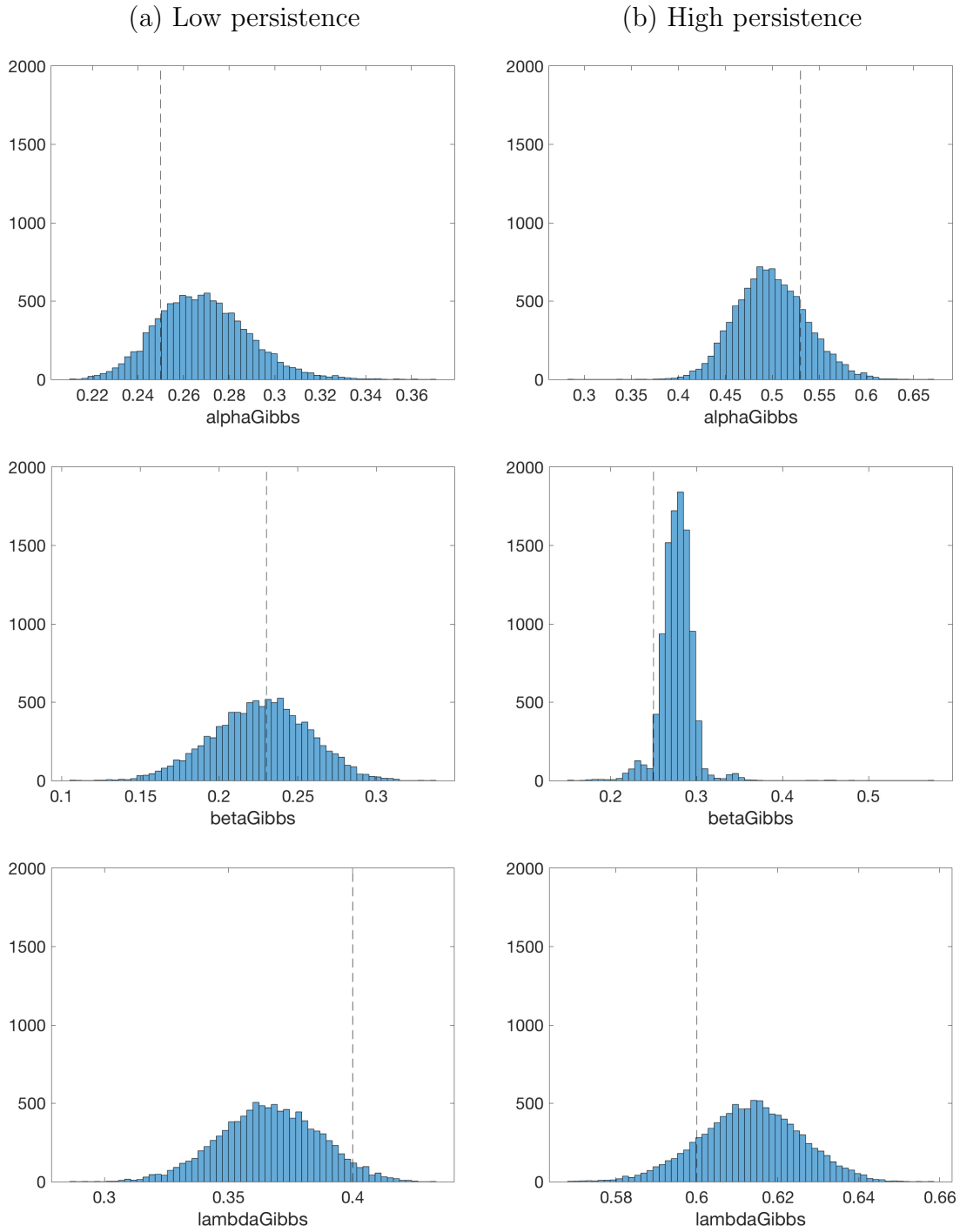


Figure 2.20: Histograms of the MCMC draws for the parameters in both settings: low persistence and high persistence, after removing the burn-in sample and thinning.

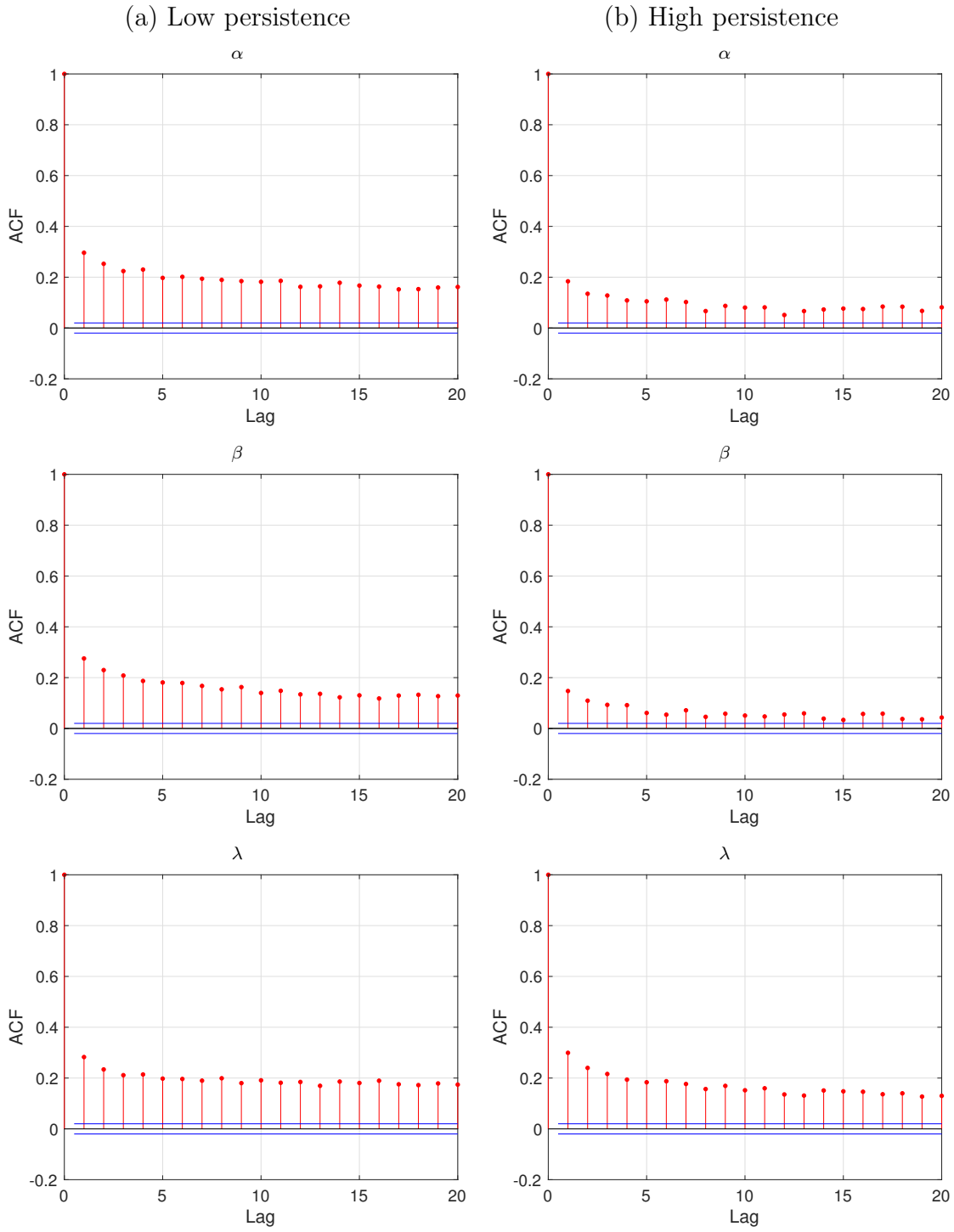


Figure 2.21: Autocorrelation function for the parameters in both low persistence and high persistence settings, after removing the burn-in sample and thinning.

Chapter 3

Dynamic Conditional Score Generalised Poisson Process [1]

Abstract. A new score driven is introduced with Generalised Poisson (GP) conditional distribution. GP is a flexible model for count data which allows for under- and over-dispersion. We provide a Bayesian inference framework and an efficient posterior approximation procedure based on Markov Chain Monte Carlo. The applications to fires show that the proposed model is well suited for capturing the over-dispersion feature of the data.

3.1 Introduction

The Generalized Poisson (GP) distribution has been introduced by Consul and Jain (1973) in order to overcome the equality of mean and variance (equidispersion) that characterizes the Poisson distribution. Consul (1989) and Consul and Famoye (2006) have studied the properties of the GP distribution. This distribution family has been used in many fields. Consul (1989) applied GP distribution to the purchasing of commodities, Tripathi et al. (1986) to profit maximization in textile manufacturing industry, Famoye (2015) and Zamani et al. (2016) to health care utilization, Wang and Famoye (1997) to households fertility decisions and (see, e.g. Consul, 1989; Famoye and Consul, 1995; Famoye et al., 2004) to modelling the number of accidents.

In time series analysis various dynamic models with GP distributed innovations have been proposed to account for over-and under-dispersion. Alzaid and Al-Osh (1993) introduced the integer-valued ARMA (INARMA) process with GP marginal distribution while Al-Nachawati et al. (1997) provided estimation method for the parameters of the GP-INAR(1). Yang et al. (2019) defined a new GP-INAR process by using a GP thinning operator. Finally, a model able

[1] In collaboration with Dario Palumbo (Ca' Foscari University of Venice, Italy and University of Cambridge, UK).

to deal with the heteroskedastic feature of count data using GP conditionally distributed observations, the GP-INGARCH process, has been introduced by Zhu (2012) while Chen and Lee (2016) introduced a zero-inflated ZIGP-GARCH version. We propose a new flexible dynamic GP model based on Dynamic Conditional Score (DCS).

DCS models are observation-driven models introduced by Creal et al. (2011, 2013) as Generalized Autoregressive Score models and studied in Harvey (2013). In DCS models the time-varying parameters are driven by the score function of the predictive density. DCS models have several advantages. First Creal et al. (2013) show that by an appropriate scaling of the score function other observation-driven models can be recovered as special cases, such as GARCH models, Autoregressive Conditional Duration (ACD) models and Autoregressive Conditional Intensity models (ACI). Second, the likelihood function is available in closed form. Lastly, structural components such as trend, seasonality and cycles can be easily included in the formulation of the parameter dynamics (see Harvey, 2007, 2013).
Application of these models can be found for example in credit risk modelling (Lucas et al., 2017), stock volatility (Koopman et al., 2018; Ayala and Blazsek, 2018), modelling of high frequency data (De Lira Salvatierra and Patton, 2015; Buccheri et al., 2020). Koopman et al. (2018) proposed Skellam GAS model for high-frequency stock price changes. Blasques et al. (2019) introduced a zero-inflated Negative Binomial GAS model for trade durations. In this chapter we provide an original application to climate change data. Climate change has been a key risk factor, affecting wildfire events in many regions of the world. Wildfires have a tremendous impact on the environment and human society in different directions. They are essential drivers of the inter-annual variability in the atmospheric growth rate of carbon dioxide and contribute significantly to the fine particulate emissions from biomass burning. Fires affect regional temperature, clouds, precipitation, and regional to large-scale atmospheric circulation patterns. See Hantson et al. (2016) for a discussion of the effects of fire occurrences and the challenges in fire-regime modeling. The impacts of fire-related pollution on human life can be deleterious by increasing health risks related to respiratory and cardiovascular conditions (e.g. see Buchholz et al., 2022; Woo et al., 2020). Wildfires put biodiversity at risk via the removal of vegetation, change habitat conditions, destroy food sources and increasing vulnerability of animals (e.g. see Pastro et al., 2011). In addition, fire, vegetation, and climate are intimately connected: climate changes modify vegetation that provides fuels for fire, and fire alters vegetation structure and composition with further feedback on climate change Hantson et al. (2016). Thus, designing fire risk policies and fire planning is fundamental for mitigating the consequences of wildfires on humans and the environment. Effective risk management requires a deep knowledge of the features of wildfires. This paper provides a simple and flexible statistical approach to fire-regime modelling and focuses only on the fire frequency at a spatially

aggregated level. Since the focus is on the characteristics of fire dynamics, the relationships with climate, vegetation, burnt area, and other aspects of fires are not considered. Addressing feedback effects is not in the scope of this research, requiring a much more elaborated approach, including dynamic vegetation and terrestrial ecosystem models. Our score-driven models do not require any knowledge of the vegetation and ecosystem features of the region considered and can be used to identify and predict periodic, long-term, and idiosyncratic components. The analysis of the components can help achieve better risk management and a more efficient long- and short-term allocation of fire-fighting resources. The abundance of remote sensed data allows for the analysis of key features of wildfires such as the frequency and the impact on society and ecosystem (Jin et al., 2015; Gebert et al., 2007; Gebert and Black, 2012; Houtman et al., 2013).

3.2 A Dynamic Generalized Poisson Model

Let $G(\theta, \boldsymbol{\psi})$, $\theta \in \Theta$, $\boldsymbol{\psi} \in \Psi$ be a parametric family of distributions where θ is a vector of parameters subject to temporal variation and ψ a parameter vector which is stable over time. Under the DCS framework the conditional distribution of the variable y_t with values in $\mathcal{Y}$, given all the past observations up to $t-1$, is $G\left(\theta_{t|t-1}, \boldsymbol{\psi}\right)$, that is

$$y_t \mid \mathcal{F}_{t-1} \sim G\left(\theta_{t|t-1}, \boldsymbol{\psi}\right), \qquad t = 1, \ldots, T, \tag{3.1}$$

where $\mathcal{F}_t = \sigma(\{y_u\}_{u \leq t})$ is the sigma-algebra generated by the past values of the process, $\boldsymbol{\psi}$ is a vector of time invariant hyperparameters, $\theta_{t|t-1}$,$t = 1, \ldots, T$ is a sequence of time-varying parameters with initial condition $\theta_{1|0}$, such that $\theta_{t|t-1} \in \mathcal{F}_{t-1}$. In the most simple DCS model the dynamics of the parameters is modelled as a first order Quasi-ARMA specification (Harvey, 2013) that can be represented as

$$\theta_{t+1|t} = (1 - \phi)\,\omega + \phi\theta_{t|t-1} + \kappa u_t, \qquad t = 1, \ldots, T \tag{3.2}$$

where ϕ and κ are constant parameters to be estimated and u_t is defined as the score of the conditional density with respect to the dynamic parameters $\theta_{t|t-1}$ evaluated at y_t

$$u_t = \frac{\partial \log f_t\left(y_t; \theta_{t|t-1},\ \boldsymbol{\psi}\right)}{\partial \theta_{t|t-1}}, \qquad t = 1, \ldots, T \tag{3.3}$$

It can sometimes be beneficial to standardise the score, a natural choice for the scaling matrix is the information quantity $\mathcal{I}_{\theta\theta}$ that gives a measure of the variance of the score (Creal et al., 2013). The important feature of this framework is that dynamics can be introduced in other elements of the time invariant parameter vector evaluating their conditional score and assuming a Quasi-ARMA specification.

3.2.1 Scores for the Generalized Poisson distribution

For the current study we will consider the modeling of data series y_t conditionally distributed with a Generalized Poisson distribution. Following the parametrization in Consul and Jain (1973), a non-negative random variable y_t GP distributed with parameters ν and λ has probability mass function (pmf)

$$f_t\left(y_t;\nu,\lambda\right)=\begin{cases}\frac{\nu(\nu+y_t\lambda)^{y_t-1}}{y_t!}e^{-\nu-y_t\lambda}, & y_t=0,1,2,\ldots\\ 0, & \text{for } y_t>m \text{ if } \lambda<0\end{cases} \tag{3.4}$$

where $\nu>0$, $\max\left(-1,-\nu/m\right)\leq\lambda\leq 1$, with m defined as the largest positive integer for which $\nu+m\lambda>0$ when $\lambda<0$. To avoid empirical difficulties and make the model more tractable, we take $0<\lambda<1$. The GP reduces to a Poisson distribution when $\lambda\to 0$ at the limit. Following standard practice in time series modelling location and scale reparametrization is applied. Since for a GP variable $E\left[y_t\right]=\nu(1-\lambda)^{-1}$ and $Var\left(y_t\right)=\nu(1-\lambda)^{-3}$, we set the location parameter $\mu=\nu(1-\lambda)^{-1}$ and the scale parameter $\varphi=(1-\lambda)^{-1}$ we have that $E\left[y_t\right]=\mu$ and $Var\left(y_t\right)=\mu\varphi^2$. The GP pmf, f_t, in the location and scale parametrization becomes

$$f_t\left(y_t;\mu,\varphi\right)=\frac{\mu}{\varphi}\frac{1}{y_t!}\left(\frac{\mu}{\varphi}+y_t\frac{(\varphi-1)}{\varphi}\right)^{y_t-1}e^{-\frac{\mu+y_t(\varphi-1)}{\varphi}},\qquad y_t=0,1,2,\ldots \tag{3.5}$$

where $\mu>0$ and $\varphi>1$.

Proposition 1. *The full information matrix with respect to the parameters* $(\mu,\varphi)'$ *is*

$$\mathcal{I}=\begin{pmatrix}\mathcal{I}_{\mu\mu} & \mathcal{I}_{\mu\varphi}\\ \mathcal{I}_{\varphi\mu} & \mathcal{I}_{\varphi\varphi}\end{pmatrix}=\begin{pmatrix}\frac{\mu+2\varphi(\varphi-1)}{\mu\varphi^2(\mu+2(\varphi-1))} & \frac{2(\varphi-1)}{\varphi^2(\mu+2(\varphi-1))}\\ \frac{2(\varphi-1)}{\varphi^2(\mu+2(\varphi-1))} & \frac{2\mu}{\varphi^2(\mu+2(\varphi-1))}\end{pmatrix} \tag{3.6}$$

See Appendix 3.6.1 for a proof. In order to have a strictly positive μ and φ larger than 1 we introduce two link functions of the form $\mu=e^{\theta_1}$ and $\varphi=1+e^{\theta_2}$. The information matrix in the new parametrization is given in the following.

Proposition 2. *The information matrix of* θ_1 *and* θ_2 *is*

$$\mathcal{I}=\begin{pmatrix}\mathcal{I}_{\theta_1\theta_1} & \mathcal{I}_{\theta_1\theta_2}\\ \mathcal{I}_{\theta_2\theta_1} & \mathcal{I}_{\theta_2\theta_2}\end{pmatrix}$$

with

$$\mathcal{I}_{\theta_1\theta_1} = \mathcal{I}_{\mu\mu}\mu^2=\frac{\mu\left(\mu+2\varphi\left(\varphi-1\right)\right)}{\varphi^2\left(\mu+2\left(\varphi-1\right)\right)} \tag{3.7}$$

$$\mathcal{I}_{\theta_2\theta_2} = \mathcal{I}_{\varphi\varphi}\left(\varphi-1\right)^2=\frac{2\mu\left(\varphi-1\right)^2}{\varphi^2\left(\mu+2\left(\varphi-1\right)\right)} \tag{3.8}$$

See Appendix 3.6.1 for a proof.

In the following, we provide two specification of the model. In the application, our specification strategies relies on the analysis of the residuals of the model. Hence, we analyse the fitted score as in (Harvey, 2013, Chapter 2) to decide which component should be modelled.

3.2.2 Dynamic Location

Following the re-parametrization introduced in the previous section: $\log \mu_{t|t-1} = \theta_{1,t|t-1}$. In order to fit the time series we are proposing the following specification for the $\theta_{1,t|t-1}$ dynamics of the GP-DCS model.

$$\begin{aligned} y_t \mid \mathcal{F}_{t-1} &\sim \mathrm{GP}\left(\mu_{t|t-1}; \varphi\right), \qquad \mu_{t|t-1} = e^{\theta_{1,t|t-1}} \\ \theta_{1,t|t-1} &= \alpha_{1,t|t-1} + \gamma_{1,t|t-1} + \xi_{1,t|t-1}, \qquad t = 1, \ldots, T \end{aligned}$$

$\alpha_{1,t|t-1}$ is a non stationary trend component that captures the persistent feature of the data with dynamics, $\gamma_{1,t|t-1}$ is a dynamic seasonal component and $\xi_{1,t|t-1}$ is a cyclical component, modelled as in Harvey (2013). The component dynamics is modelled as follows.

- Trend component:

$$\alpha_{1,t+1|t} = (1 - \phi_1)\,\omega_1 + \phi_1 \alpha_{1,t|t-1} + \kappa_{1,\alpha} u_{1,t}, \qquad t = 1, \ldots, T \tag{3.9}$$

  where $u_{1,t} = \frac{\partial \log f_t}{\partial \theta_1} \mathcal{I}^{-1}_{\theta_1 \theta_1}$ be the standardized score with respect to θ_1 and $\phi_1 \in (-1, 1)$.

- Seasonal component:

$$\begin{aligned} \gamma_{1,t|t-1} &= \boldsymbol{z}'_{1,t} \boldsymbol{\gamma}_{1,t|t-1} \\ \boldsymbol{\gamma}_{1,t+1|t} &= \boldsymbol{\gamma}_{1,t|t-1} + \boldsymbol{g}(s_1, t) \kappa_{1,\gamma} u_{1,t}, \end{aligned} \tag{3.10}$$

  where $s_1 \in \mathbb{N}$ indicates the number of seasons, $\boldsymbol{z}_{1,t}$ is a deterministic time varying vector with value $\boldsymbol{e}_j$ with $j = 1 + (t-1) \mathrm{mod} s$ where $\boldsymbol{e}_j$ is the j-th component of the standard orthonormal basis of $\mathbb{R}^s$. The vector valued function $\boldsymbol{g}(s_1, t)$ with values in $\mathbb{R}^s$ such that the j-th element is equal to 1 if $j = 1 + (t-1) \mathrm{mod} s$ and equal to $-1/(s_1 - 1)$ otherwise. $\kappa_{1,\gamma}$ is an unknown parameter to estimate (e.g., see Harvey, 2013, Ch. 3; Harvey, 1990, Ch. 2). In this way it can be sure that the amount of each season's component at each time period t sums to 0.

- Cyclical component: $\xi_{1,t|t-1}$ where

$$\begin{aligned} \xi_{1,t|t-1} &= \rho_1 \cos(f_1) \xi_{1,t-1|t-2} + \rho_1 \sin(f_1) \xi^*_{1,t-1|t-2} + \kappa_{1,\xi} u_{1,t-1} \\ \xi^*_{1,t|t-1} &= -\rho_1 \sin(f_1) \xi_{1,t-1|t-2} + \rho_1 \cos(f_1) \xi^*_{1,t-1|t-2} + \kappa^*_{1,\xi} u_{1,t-1} \end{aligned} \tag{3.11}$$

with $j = 1, ..., [s/2]$ and $f_1 \in \{2\pi j/T,\, j = 1, \ldots, [T/2]\}$. Since $\sin(\pi) = 0$, the two processes decouple at frequency π, and the component becomes autoregressive. The component reverts to a sine-cosine wave when the corresponding $\kappa_{1,\xi}$ and $\kappa^*_{1,\xi}$ are both zero. For $\rho_1 = 1$ the cycle is non stationary. For identification purposes we assume $\kappa_{1,\xi} = \kappa^*_{1,\xi}$.

In order to initialize the filter, since both the components $\gamma_{t|t-1}$ and $\mu_{t|t-1}$ are non stationary, μ_0 and $\boldsymbol{\gamma}_{1|0}$ are treated as parameters and estimated. However, since the elements of $\boldsymbol{\gamma}_{1|0}$ have to sum up to zeros, there are only $s_1 - 1$ free parameters in $\boldsymbol{\gamma}_{1|0}$ to estimate. To estimate the model on the proposed dataset there are $s_1 = 12$ season to be modelled. This can be also confirmed by an analysis of the spectrum [FFT] of the time series which reveal a peak around period 9 and 12. Hence, the vector of the static parameters to be estimated is $\boldsymbol{\psi} = (\phi_1, \omega_1, \kappa_{1,\alpha}, \gamma_{1|0}, \kappa_{1,\gamma}, \rho_1, \kappa_{1,\xi})$.

3.2.3 Dynamic Scale

Following the re-parametrization introduced in the previous section: $\varphi_{t|t-1} = 1 + \exp(\theta_{2,t|t-1})$. In order to fit the time series we are proposing the following specification for the $\varphi_{t|t-1}$ dynamics of the GP-DCS model, while the dynamics for $\mu_{t|t-1}$ is the same as in Section 3.2.2.

$$\begin{aligned} &y_t \mid \mathcal{F}_{t-1} \sim \text{GP}\left(\mu_{t|t-1}, \varphi_{t|t-1}\right), \qquad \varphi_{t|t-1} = 1 + e^{\theta_{2,t|t-1}} \\ &\theta_{2,t|t-1} = \alpha_{2,t|t-1} + \gamma_{2,t|t-1} + \xi_{2,t|t-1}, \qquad t = 1, \ldots, T \end{aligned} \tag{3.12}$$

$\alpha_{2,t|t-1}$ is a non stationary trend component, $\gamma_{2,t|t-1}$ is a dynamic seasonal component and $\xi_{2,t|t-1}$ is a cyclical component. The components are modelled as before:

- Trend component:

$$\alpha_{2,t+1|t} = (1 - \phi_2)\,\omega_2 + \phi_2 \alpha_{2,t|t-1} + \kappa_{2,\alpha} u_{2,t}, \qquad t = 1, \ldots, T \tag{3.13}$$

  where $u_{2,t} = \frac{\partial \log f_t}{\partial \theta_2} \mathcal{I}^{-1}_{\theta_2 \theta_2}$ be standardized score with respect to θ_2 and $\phi_2 \in (-1, 1)$.

- Seasonal component:

$$\begin{aligned} \gamma_{2,t|t-1} &= \boldsymbol{z}'_{2,t} \boldsymbol{\gamma}_{2,t|t-1} \\ \boldsymbol{\gamma}_{2,t+1|t} &= \boldsymbol{\gamma}_{2,t|t-1} + \boldsymbol{g}(s_2, t) \kappa_{2,\gamma} u_{2,t}, \end{aligned} \tag{3.14}$$

  where $s_2 \in \mathbb{N}$ indicates the number of seasons, $\boldsymbol{z}_{2,t}$ is a deterministic time varying vector with value $\boldsymbol{e}_j$ with $j = 1 + (t-1) \text{mod} s$ where $\boldsymbol{e}_j$ is the j-th component of the standard orthonormal basis of $\mathbb{R}^s$. The vector valued function $\boldsymbol{g}(s_2, t)$ with values in $\mathbb{R}^s$ such that the element $j = 1 + (t-1) \text{mod} s$ equal to 1 and $-1/(s_2 - 1)$ otherwise. $\kappa_{2,\gamma}$ is an unknown parameter to estimate. In this way it can be sure that the amounts of each season's component at each time period t sums to 0.

- Cyclical component: Cyclical component: $\xi_{2,t|t-1}$ where

$$\begin{aligned}\xi_{2,t|t-1} &= \rho_2 \cos(f_2)\xi_{2,t-1|t-2} + \rho_2 \sin(f_2)\xi^*_{2,t-1|t-2} + \kappa_{2,\xi} u_{2,t-1} \\ \xi^*_{2,t|t-1} &= -\rho_2 \sin(f_2)\xi_{2,t-1|t-2} + \rho_2 \cos(f_2)\xi^*_{2,t-1|t-2} + \kappa^*_{2,\xi} u_{2,t-1},\end{aligned} \tag{3.15}$$

  with $j = 1, ..., [s/2]$ and $f_2 \in \{2\pi j/T,\ j = 1, \ldots, [T/2]\}$. Since $\sin(\pi/2) = 0$, the two processes decouple at frequency $\pi/2$, and the component becomes autoregressive. The component reverts to a sine-cosine wave when the corresponding $\kappa_{2,\xi}$ and $\kappa^*_{2,\xi}$ are both zero. For $\rho_2 = 1$ the cycle is non stationary. For identification purposes we need $\kappa_{2,\xi} = \kappa^*_{2,\xi}$.

Hence, the vector of the static parameters to be estimated is $\boldsymbol{\psi} = (\phi_1, \omega_1, \kappa_{1,\alpha}, \gamma_{1|0}, \kappa_{1,\gamma}, \rho_1, \kappa_{1,\xi}, \phi_2, \omega_2, \kappa_{2,\alpha}, \gamma_{2|0}, \kappa_{2,\gamma}, \rho_2, \kappa_{2,\xi})$.

3.2.4 Model properties

In the context of observation-driven time varying parameter models, filter invertibility is essential to obtain consistent statistical inference, (Straumann and Mikosch, 2006; Wintenberger, 2013, e.g., see). Blasques et al. (2018) lay down strict conditions for the invertibility of a score-driven model for continuous data with a smooth probability density function. We focus on a special case of our GP-DCS model, given by a first-order location score-driven model with trend component, that can be constructed as

$$y_t \mid \mathcal{F}_{t-1} \sim \text{GP}\left(\theta_{1,t|t-1}; \boldsymbol{\psi}\right), \qquad \theta_{1,t|t-1} = \log \mu_{t|t-1} \tag{3.16}$$

$$\theta_{1,t+1|t} = g(\theta_{1,t|t-1}, y_t, \boldsymbol{\psi}) \tag{3.17}$$

where $\theta_{1,t|t-1}(\boldsymbol{\psi}) = \log \mu_{t|t-1}$ is the time-varying parameter, with values in $\mathcal{F}_\psi \subset \mathbb{R}$, $u_{1,t} = \frac{\partial \log f_t}{\partial \theta_1} \mathcal{I}^{-1}_{\theta_1\theta_1}$ is the scaled score, $\boldsymbol{\psi} = (\omega_1, \phi_1, \kappa_1, \varphi)'$ is a vector of time-invariant hyperparameters to be estimated taking values in the compact set Ψ and g is a continuous function from $\mathcal{F}_\psi \times \mathcal{Y} \times \Psi$ into $\mathcal{F}_\psi$ differentiable on its first coordinate. In our dynamic-location model $g(\theta_{1,t|t-1}, y_t, \boldsymbol{\psi}) = (1 - \phi_1)\omega_1 + \phi_1\theta_{1,t|t-1} + \kappa_1 u_{1,t}$.

The filter $\{\hat{\theta}_{1,t|t-1}(\boldsymbol{\psi})\}_{t\in\mathbb{N}}$ initialized at some point $\hat{\theta}_{1,1} = \omega_1 \in \mathbb{R}$ is said to be invertible if $\hat{\theta}_{1,t|t-1}(\boldsymbol{\psi})$ converges almost surely exponentially fast (e.a.s.)[2] to a unique strictly stationary and ergodic sequence $\{\theta_{1,t|t-1}(\boldsymbol{\psi})\}_{t\in\mathbb{N}}$ that is

$$\left|\hat{\theta}_{1,t|t-1}(\boldsymbol{\psi}) - \theta_{1,t|t-1}(\boldsymbol{\psi})\right| \xrightarrow{e.a.s.} 0 \quad \text{as} \quad t \to \infty$$

Let $\Lambda_t(\boldsymbol{\psi})$ be the Lipschitz coefficient defined as

$$\Lambda_t(\boldsymbol{\psi}) := \sup_{\theta_1 \in \mathcal{F}_\psi} |x_t(\theta_1, \boldsymbol{\psi})|$$

[2] A sequence of random variables $\{w_t\}_{t\in\mathbb{N}}$ is said to converge e.a.s. to another sequence $\{\hat{w}_t\}_{t\in\mathbb{N}}$ if there exists a constant $\gamma > 1$ such that $\gamma^t|\hat{w}_t - w_t| \overset{a.s.}{\to} 0$ as t diverges.

where $x_t(\theta_1, \boldsymbol{\psi}) = \partial g(\theta_1, y_t, \boldsymbol{\psi}) / \partial \theta_1$. Following Proposition 3.1 in Blasques et al. (2018) we can establish invertibility of the score-driven filter if the following result holds

Proposition 3. *(Filter Invertibility). Let the observed data process $\{y_t\}_{t\in\mathbb{N}}$ be strictly stationary and ergodic with bounded logarithmic moments $E\left[\log^+ |y_t|\right] < \infty$. Moreover let Ψ be a compact set which ensures that*

(i) There exists a $\hat{\theta}_{1,1} = \omega_1 \in \mathcal{F}_\psi$ such that $E \log^+ \sup_{\boldsymbol{\psi}\in\Psi} \left| g\left(\hat{\theta}_{1,1}, y_t, \boldsymbol{\psi}\right)\right| < \infty$.

(ii) $E \sup_{\boldsymbol{\psi}\in\Psi} \sup_{\theta_1 \in \mathcal{F}_\psi} \log^+ |x_t(\theta_1, \boldsymbol{\psi})| < \infty$.

(iii) $\log \Lambda_1(\boldsymbol{\psi})$ is a.s. continuous on Ψ and $E \log \Lambda_1(\boldsymbol{\psi}) < 0$ for any $\boldsymbol{\psi} \in \Psi$.

Then the filter $\{\hat{\theta}_{1,t|t-1}(\boldsymbol{\psi})\}_{t\in\mathbb{N}}$ is invertible, uniformly in $\boldsymbol{\psi} \in \Psi$.

See Appendix 3.6.1 for a proof of the proposition.

3.3 Bayesian Inference

We propose a Bayesian approach to inference for our Generalized Poisson DCS model which allows us to include extra-sample information about the model and the parameter through the prior choice. Bayesian estimation comes naturally with a quantification of the parameter and model uncertainty.

3.3.1 Prior Distributions

For the dynamic location model in Section 3.2.2 we propose the following prior distributions. A uniform distribution for ϕ_j, $j = 1, 2$, $\phi_j \sim \mathcal{U}(a_\phi, b_\phi)$, and ρ_j, $\rho_j \sim \mathcal{U}(a_\rho, b_\rho)$. The proposal of $a_\phi = -1$ and $b_\phi = 1$ comes from the stationarity assumptions of the trend component, while we set $a_\rho = 0$ and $b_\rho = 1$ from the assumptions of the model.

For the other parameters we propose a Normal prior with location-scale parametrization $\mathcal{N}(m, s)$. Setting $m_\omega = 0$ e $s_\omega = 1$, ω_j, $j = 1, 2$, takes values in the interval $[-1.64, 1.64]$ with the 95% of probability, which seems a reasonable assumption in our applications. Despite of the informative-prior assumption, in our example the posterior concentrates on values far from zero. We set $m_\kappa = 0$ and $s_\kappa = 2$ with κ_i, $i = \alpha_1, \gamma_{1,2}, \xi_{1,2}$, taking values in the interval $[-3.92, 3.92]$ with the 95% of probability, which seems a reasonable assumption in our applications. We set $m_{\kappa,2} = 0$ e $s_{\kappa,2} = 0.05$, for $\kappa_{\alpha,2}$ which takes values in the interval $[-0.098, 0.098]$ with the 95% of probability. For automatic use, we recommend this default setting which tends to avoid extremes in the time-varying parameters. Finally, setting $m_\gamma = 0$ e $s_\gamma = 2$, γ_i with $i = 1, 2$, takes values in the interval $[-1.64, 1.64]$ with the 95% of probability.

3.3.2 Posterior Approximation

Let $\mathbf{y} = (y_1, \ldots, y_T)'$ be the collection of observations and $\boldsymbol{\psi}$ is the parameter vector, the likelihood associated to the model in Eq. 3.12 is

$$L(\boldsymbol{y}|\boldsymbol{\psi}) = \prod_{t=1}^{T} \frac{\mu_t}{\varphi_t} \frac{1}{y_t!} \left(\frac{\mu_t}{\varphi_t} + y_t \frac{(\varphi_t - 1)}{\varphi_t} \right)^{y_t - 1} e^{-\frac{\mu_t + y_t(\varphi_t - 1)}{\varphi_t}}. \tag{3.18}$$

For the dynamic location model in 3.2.2, $\boldsymbol{\psi} = (\theta_2, \phi_1, \omega_1, \kappa_{1,\alpha}, \kappa_{1,\gamma}, \rho_1, \kappa_{1,\psi})'$. For the dynamic scale-location model in 3.2.3, $\boldsymbol{\psi} = (\phi_1, \omega_1, \kappa_{1,\alpha}, \kappa_{1,\gamma}, \rho_1, \kappa_{1,\psi}, \phi_2, \omega_2, \kappa_{2,\alpha}, \kappa_{2,\gamma}, \rho_2, \kappa_{2,\psi})'$. Let $\pi(\boldsymbol{\psi})$ be the joint prior distribution defined in the previous section, the joint posterior distribution $\pi(\boldsymbol{\psi}, \boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\psi})\pi(\boldsymbol{\psi})$ is not tractable, but efficient Monte Carlo Markov Chain (MCMC) approximation methods can be applied. In this chapter , we consider the adaptive random scan adaptive Metropolis-within-Gibbs studied in Łatuszyński et al. (2013). Let $(\Psi, \mathcal{B}(\Psi))$ be a d-dimensional space, such that $\Psi = \Psi_1 \times \ldots \times \Psi_d$ and write $\boldsymbol{\psi}^{(n)} \in \Psi$ as $\boldsymbol{\psi}^{(n)} = (\boldsymbol{\psi}_1^{(n)}, \ldots, \boldsymbol{\psi}_d^{(n)})$. Let $\boldsymbol{\psi}_{-i}^{(n)} = (\boldsymbol{\psi}_1^{(n)}, \ldots, \boldsymbol{\psi}_{i-1}^{(n)}, \boldsymbol{\psi}_{i+1}^{(n)}, \ldots, \boldsymbol{\psi}_d^{(n)})$ be the parameter vector obtained dropping the i-th component from $\boldsymbol{\psi}^{(n)}$ with values in $\Psi_{-i} = \Psi_1 \times \ldots \times \Psi_{i-1} \times \Psi_{i+1} \times \ldots \times \Psi_d$ and with $\pi(\cdot|\boldsymbol{\psi}_{-i})$ the conditional distribution of $\boldsymbol{\psi}_i$ given $\boldsymbol{\psi}_{-i}$. The adaptive random scan adaptive Metropolis within Gibbs, draws $\boldsymbol{\psi}^{(n)}$ given $\boldsymbol{\psi}^{(n-1)}$ performing a Metropolis Hastings step, by first choosing coordinates at random according to some selection probabilities $\alpha = (\alpha_1, \ldots, \alpha_d)$. Therefore, given $\boldsymbol{\psi}_{-i}^{(n-1)}$ the i-th coordinate is selected with probability α_i and $\boldsymbol{\psi}_i^{(n-1)}$ is updated by drawing $\boldsymbol{\psi}^*$ from the proposal distribution $Q_{\boldsymbol{\psi}_{-i}^{(n-1)}, v_i^{(n-1)}}(\boldsymbol{\psi}_i^{(n-1)}, \cdot)$. The proposal is chosen adaptively from the distribution family $Q_{\Psi_{-1}, v}$ by setting the parameters to $\gamma_i^{(n)}$. The sampler iterates the following steps

1. Set $\alpha^{(n)} = R_n(\alpha^{(0)}, \ldots, \alpha^{(n-1)}, \boldsymbol{\psi}^{(n-1)}, \ldots, \boldsymbol{\psi}^{(0)}, v^{(n-1)}, \ldots, v^{(0)}) \in A$.
2. Set $v^{(n)} = R'_n(\alpha^{(0)}, \ldots, \alpha^{(n-1)}, \boldsymbol{\psi}^{(n-1)}, \ldots, \boldsymbol{\psi}^{(0)}, v^{(n-1)}, \ldots, v^{(0)}) \in G_1 \times \cdots \times G_n$.
3. Choose coordinate $i \in 1, \ldots, d$ according to selection probabilities α_n
4. Draw $\boldsymbol{\psi}_i^* \sim Q_{\boldsymbol{\psi}_{-i}^{(n-1)}, v_i^{(n-1)}}(\boldsymbol{\psi}_i^{(n-1)}, \cdot)$.
5. With probability

$$\rho_i^{(n)} = \min \left(1, \frac{\pi(\boldsymbol{\psi}_i^*|\boldsymbol{\psi}_{-i}^{(n-1)}) Q_{\boldsymbol{\psi}_{-i}^{(n-1)}, v_i^{(n-1)}}(\boldsymbol{\psi}_i^*|\boldsymbol{\psi}_i^{(n-1)})}{\pi(\boldsymbol{\psi}_i^{(n-1)}|\boldsymbol{\psi}_{-i}^{(n-1)}) Q_{\boldsymbol{\psi}_{-i}^{(n-1)}, v_i^{(n-1)}}(\boldsymbol{\psi}_i^{(n-1)}, \boldsymbol{\psi}_i^*)} \right) \tag{3.19}$$

accept the proposal and set $\psi^{(n)} = (\boldsymbol{\psi}_1^{(n-1)}, \ldots, \boldsymbol{\psi}_{i-1}^{(n-1)}, \boldsymbol{\psi}^*, \boldsymbol{\psi}_{i+1}^{(n-1)}, \ldots, \boldsymbol{\psi}_d^{(n-1)})$ otherwise reject and set $\boldsymbol{\psi}^{(n)} = \boldsymbol{\psi}^{(n-1)}$.

The adaptive proposal distribution is chosen following Andrieu and Thoms (2008). In particular, we assume the distribution $Q_{\boldsymbol{\psi}_{-i},v_i}(\boldsymbol{\psi}_i,\cdot)$ has parameters $v_i = \{g_i, l_i, \mathbf{m}_i, S_i\}$ which update as follows:

$$\log(l_i^{(n+1)}) = \log(l_i^{(n)}) + g_i^{(n+1)}(\rho_i^{(n+1)} - \bar{\rho}) \quad (3.20)$$

$$\mathbf{m}_i^{(n+1)} = \mathbf{m}_i^{(n)} + g_i^{(n+1)}(\boldsymbol{\psi}_i^{(n+1)} - \mathbf{m}_i^{(n)}) \quad (3.21)$$

$$S_i^{(n+1)} = S_i^{(n)} + g_i^{(n+1)}((\boldsymbol{\psi}_i^{(n+1)} - \mathbf{m}_i^{(n)})(\boldsymbol{\psi}_i^{(n+1)} - \mathbf{m}_i^{(n)})' - S_i^{(n)}) \quad (3.22)$$

where $g_i^{(n+1)} = (n+1)^{-a_i}$ is the adaptive scale for the i-th parameter, $\bar{\rho}$ is the expected acceptance probability. Following the suggestions in Roberts et al. (1997) and Andrieu and Thoms (2008) we choose $\bar{\rho} = 0.44$.

3.4 Real Data Application

3.4.1 Fires Dataset

We focus on fires in Brazil collected by INPE (Instituto Nacional de Pesquisas Espaciais) at monthly frequency from June 1998 to September 2020 (see top plot in Figure 3.1). There are used all ten satellites having optical sensors operating in the average thermal range of 4um and which INPE is able to receive. At September 2020, polar satellite images are operationally processed in the Imaging Division, DGI, and Satellite and Environmental Systems Division, DSA, such as the AVHRR-NOAA (Ressl et al., 2009; Chuvieco and Martin, 1994; Kidwell, 1991), the NASA's Moderate Resolution Imaging Spectroradiometer (MODIS), which is a remotely sensed fire datasets (Giglio et al., 2016, 2003) providing observations of fires over the entire globe, the VIIRS (Li et al., 2018; Csiszar et al., 2016) and the images of the geostationary satellites, VA-16 and MSG-3.
Each polar-orbiting satellite produces at least two sets of images per day and the geostationary generates six images per hour, with a total of INPE automatically processing over 200 images per day specifically to detect vegetation burning points. Receptions are held at the stations of Cachoeira Paulista, SP (near the border with RJ) and Cuiab, MT. INPE monitors fires through the use of satellite images therefore they can collect the number of fires but not their extent.
The series is stationary (see ADF test in Table 3.1), exhibits over-dispersion (see index of dispersion in Table 3.1), autocorrelation and seasonality (see mid and bottom panel in Figure 3.1). Hence, the decision to fit our GP-DCS model with trend, seasonal and cyclical components in the dynamics of the location. Since the square of the number of fires exhibits autocorrelation and periodicity (see Figure 3.2) additional trend and cyclical components are included in the dynamics of the scale.
The estimation results are reported in Table 3.2. One AR(1) component in both dynamic location and dynamic scale is sufficient to significantly reduce the residual autocorrelation. There is evidence of large persistence in the scale ($\hat{\phi}_2 = 0.932$) and in the location ($\hat{\phi}_1 = 0.797$). The evidence of an upward trend,

calls for the adoption long-term planning strategies to reduce wildfire events and their impact on climate change.
As regards the seasonal components, the spectrum of the series shows clear picks at 0.527 (twelve months), 1.0431 (six months), 1.5708 (four months) and 2.61 (two months). Thus, an yearly cycle component and twelve monthly seasonal components are included. There is evidence of the reduced number of fires during the less hot months of the year ($(-1.277 \leq \hat{\gamma}_j \leq -0.004$ with $j \in (1, 7, 8, 9, 10, 11, 12)$) and increased number in the hottest months with a peak in September ($\hat{\gamma}_4 = 1.71$) consistently with the dynamics of the maximum temperature levels in Brazil. This evidence can be used for an efficient allocation of fire-fighting resources, for example by increasing the number of fire-fighters during the months with an higher expected number of wildfires.

Table 3.1, shows additional information about the INPE dataset.

	Mean	Variance	Skewness	Kurtosis	Idx Disp.	ADF	PP
Fires	1.87e+04	5.7e+08	2.098	7.992	86.215	-8.828**	-7.327**

Table 3.1: Descriptive statistics, stationary tests and normality test for the fires datasets. ADF is the Augmented Dickey-Fuller test with stationarity as null hypothesis, PP is the Phillips-Perron test with stationarity as null hypothesis. The symbol ** means that the null hypothesis is rejected at 5% significance level.

Parameters	Mean	Std	CI
ω_1	8.599	0.026	$(8.549, 8.637)$
$\kappa_{1,\alpha}$	0.561	0.003	$(0.555, 0.567)$
ϕ_1	0.797	0.003	$(0.788, 0.800)$
$\kappa_{1,\gamma}$	0.242	0.001	$(0.240, 0.244)$
$\gamma_{0,1}$	-0.004	0.002	$(-0.008, -0.001)$
$\gamma_{0,2}$	0.233	0.011	$(0.206, 0.247)$
$\gamma_{0,3}$	1.570	0.017	$(1.539, 1.610)$
$\gamma_{0,4}$	1.710	0.014	$(1.686, 1.741)$
$\gamma_{0,5}$	1.209	0.009	$(1.193, 1.222)$
$\gamma_{0,6}$	0.323	0.007	$(0.309, 0.333)$
$\gamma_{0,7}$	-0.034	0.005	$(-0.043, -0.024)$
$\gamma_{0,8}$	-1.105	0.010	$(-1.124, -1.085)$
$\gamma_{0,9}$	-0.908	0.007	$(-0.927, -0.897)$
$\gamma_{0,10}$	-1.223	0.010	$(-1.239, -1.202)$
$\gamma_{0,11}$	-1.277	0.017	$(-1.304, -1.233)$
$\gamma_{0,12}$	-0.493	0.010	$(-0.516, -0.473)$
$\kappa_{1,\xi}$	0.155	0.022	$(0.126, 0.209)$
ρ_1	0.957	0.052	$(0.794, 0.999)$
ω_2	3.386	0.027	$(3.330, 3.431)$
ϕ_2	0.932	0.049	$(0.786, 0.984)$
$\kappa_{2,\alpha}$	0.047	0.009	$(0.028, 0.067)$
$\kappa_{2,\xi}$	0.070	0.013	$(0.051, 0.092)$
ρ_2	0.997	0.002	$(0.993, 1.000)$

Table 3.2: GP-DCS parameter estimates. Posterior mean (Mean), standard deviation (Std) and 95% credible intervals (CI) of the parameters.

In Figure 3.3 we show the estimated components of the model. The cycle component is less evident int he first part of the series, while it becomes more pronounced in the second part due to the effect of climate change. We finally report the estimated dynamic parameters of the Generalised Poisson distribution, ν and λ, and the estimated over-dispersion in the data in Figure 3.4. From the figure, we can notice that the overdispersion heavily increase when the series takes large values.

3.5 Conclusions

We introduce a new family of stochastic processes with values in the set of the integers. The data series follows a generalized Poisson distribution with time-varying parameters. We assume a DCS dynamics and we derived the scores of the process. We provide a Bayesian inference procedure and an efficient Monte Carlo Markov Chain sampler for posterior approximation. Inference fires data show that the proposed DCS model is well suited for capturing persistence in the conditional moments and in the over-dispersion feature of the data.

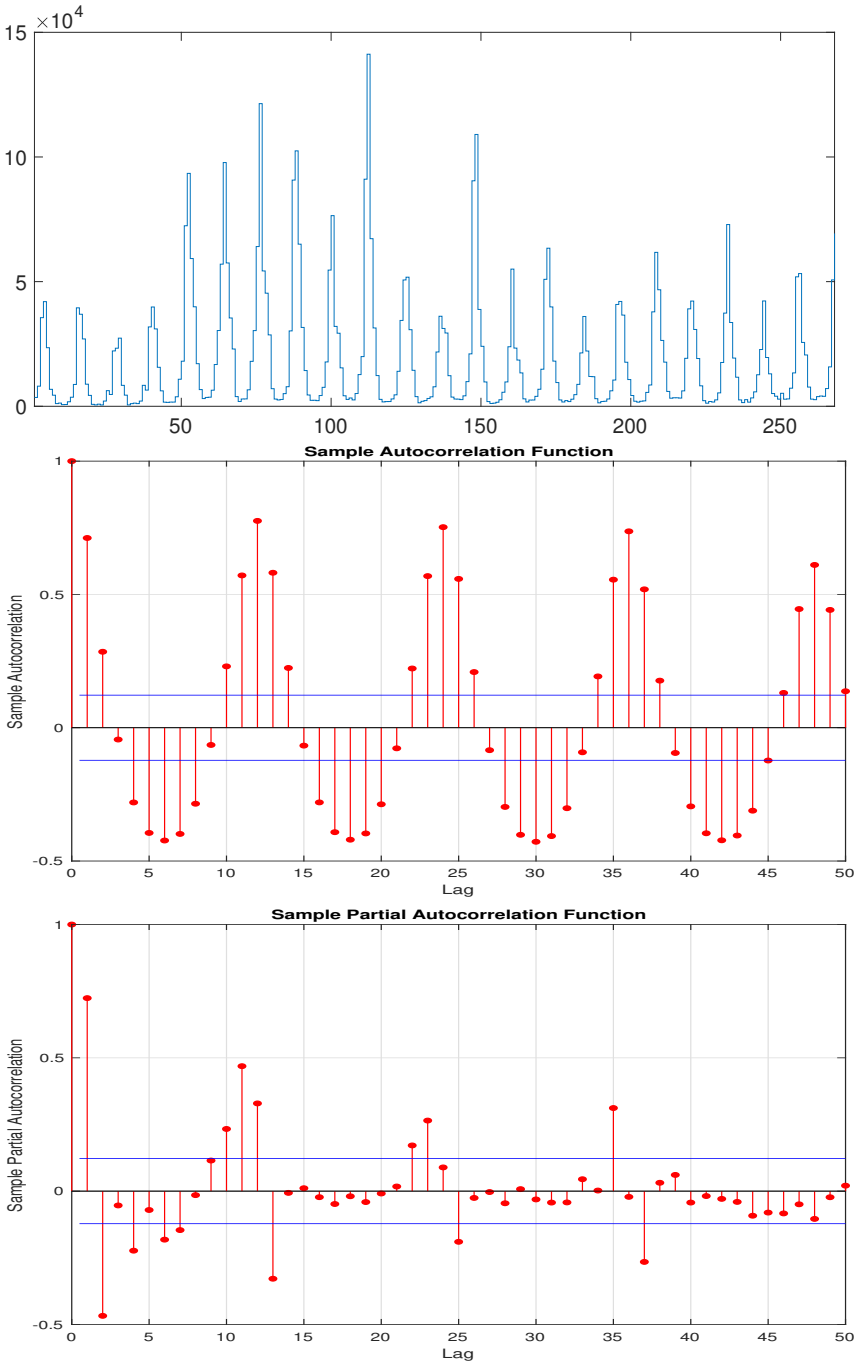


Figure 3.1: Top: total number of fires at a monthly frequency from June 1998 to September 2020 in Brazil. Middle: autocorrelation of the series. Bottom: partial autocorrelation of the series.

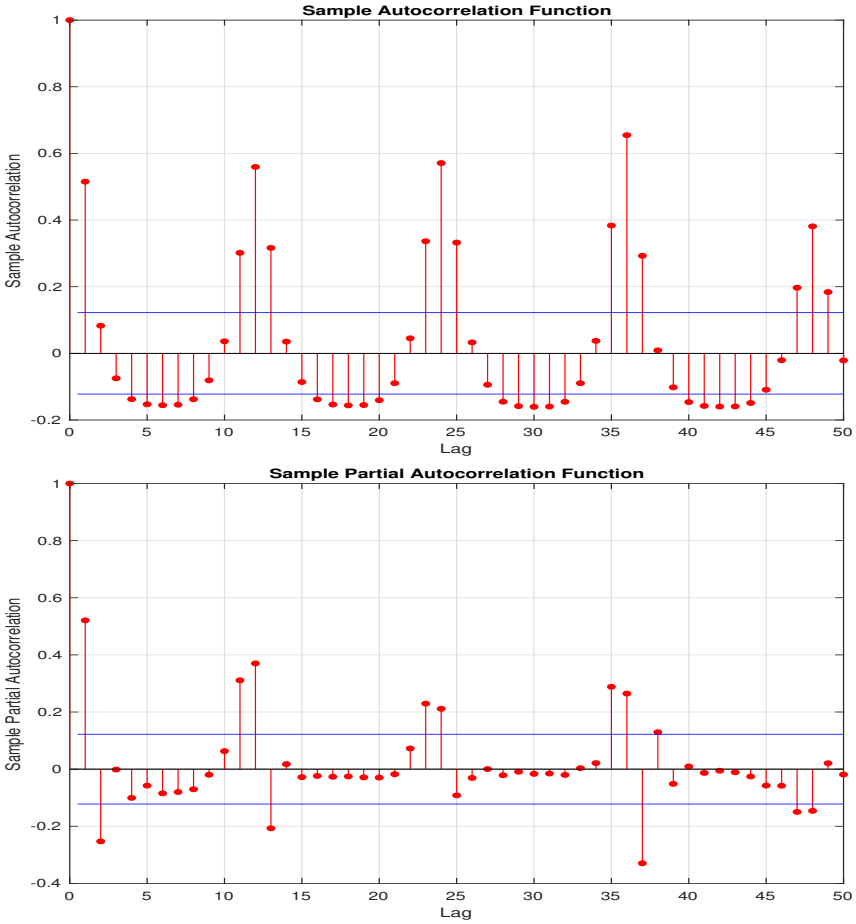


Figure 3.2: Top: autocorrelation of the squared number of fires. Bottom: Partial autocorrelation of the squared number of fires.

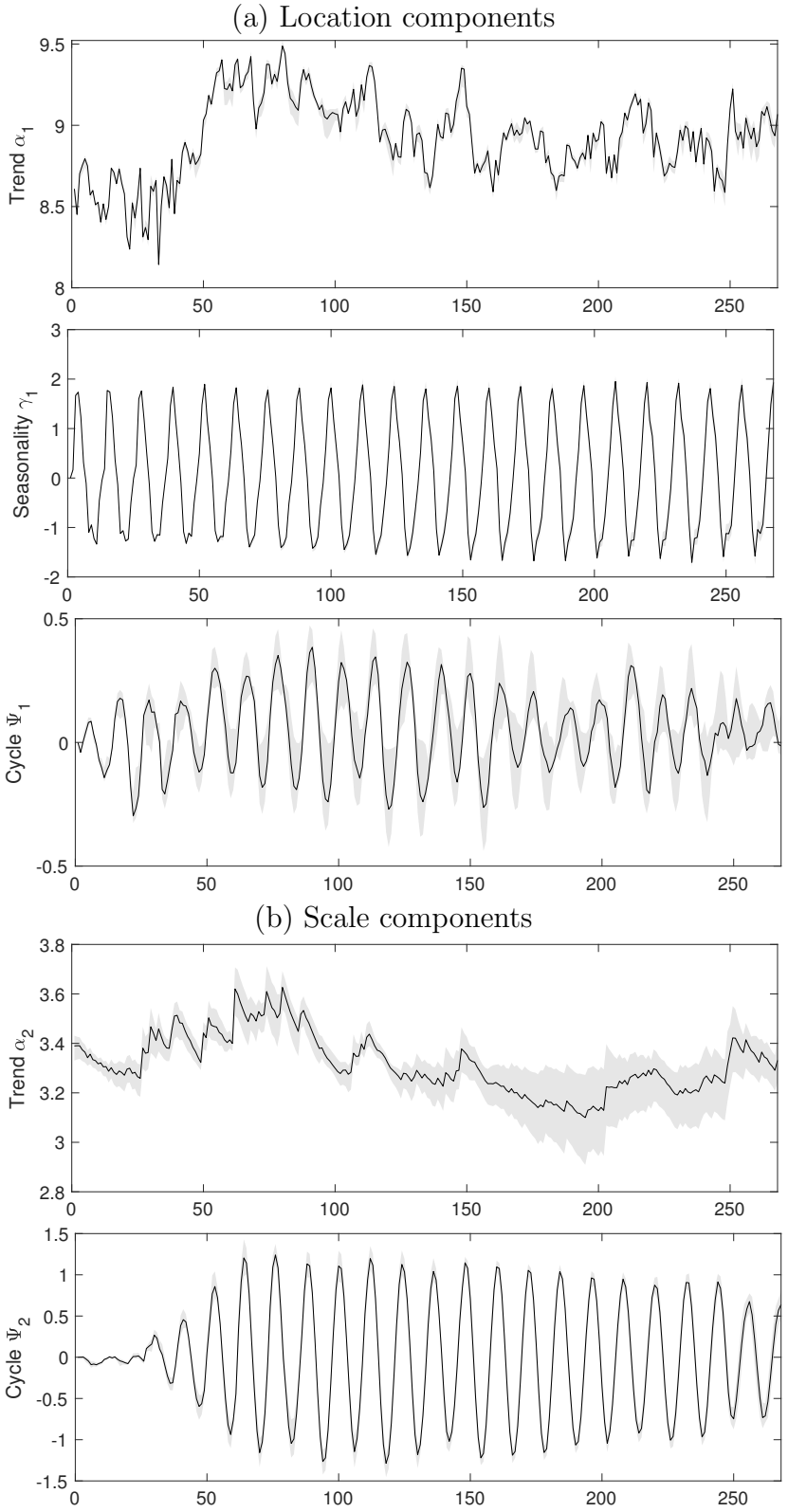


Figure 3.3: Estimated components for the location (a) and scale (b) of the GP-DCS model. In each plot the median (solid line) and 95% credible intervals (grey area).

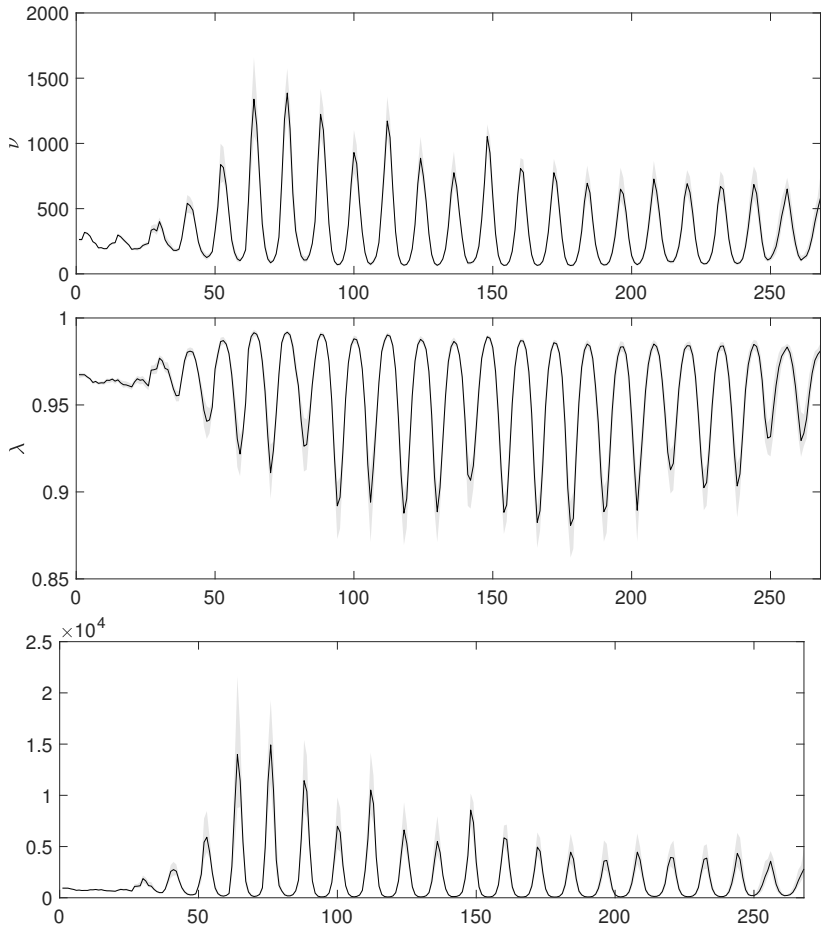


Figure 3.4: Estimated parameters of the GP-DCS model (top and mid figures) and estimated overdispersion in the data (bottom figure). In each plot the median (solid line) and 95% credible intervals (grey area).

Bibliography

Al-Nachawati, H., Alwasel, I., and Alzaid, A. (1997). Estimating the parameters of the generalized Poisson AR(1) process. *Journal of Statistical Computation and Simulation*, 56(4):337–352.

Alzaid, A. A. and Al-Osh, M. A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics*, 45(2):223–232.

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and computing*, 18(4):343–373.

Ayala, A. and Blazsek, S. (2018). Equity market neutral hedge funds and the stock market: An application of score-driven copula models. *Applied Economics*, 50(37):4005–4023.

Blasques, F., Gorgi, P., Koopman, S. J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, 12(1):1019 – 1052.

Blasques, F., Holỳ, V., and Tomanová, P. (2019). Zero-Inflated Autoregressive Conditional Duration Model for Discrete Trade Durations with Excessive Zeros. Technical report, Tinbergen Institute Discussion Paper.

Buccheri, G., Bormetti, G., Corsi, F., and Lillo, F. (2020). A score-driven conditional correlation model for noisy and asynchronous data: An application to high-frequency covariance dynamics. *Journal of Business & Economic Statistics*, pages 1–17.

Buchholz, R. R., Park, M., Worden, H. M., Tang, W., Edwards, D. P., Gaubert, B., Deeter, M. N., Sullivan, T., Ru, M., Chin, M., et al. (2022). New seasonal pattern of pollution emerges from changing north american wildfires. *Nature communications*, 13(1):1–9.

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

Chuvieco, E. and Martin, M. P. (1994). Global fire mapping and fire danger estimation using avhrr images. *Photogrammetric Engineering and Remote Sensing*, 60(5):563–570.

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.

Creal, D., Koopman, S. J., and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics*, 29(4):552–563.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Csiszar, I., Schroeder, W., Giglio, L., Mikles, V., and Tsidulko, M. (2016). The NOAA NDE active fire EDR external users manual, edited.

De Lira Salvatierra, I. and Patton, A. J. (2015). Dynamic copula models and high frequency data. *Journal of Empirical Finance*, 30:120–135.

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):127–138.

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*, 2:287–295.

Gebert, K. M. and Black, A. E. (2012). Effect of suppression strategies on federal wildland fire expenditures. *Journal of Forestry*, 110(2):65–73.

Gebert, K. M., Calkin, D. E., and Yoder, J. (2007). Estimating suppression expenditures for individual large wildland fires. *Western Journal of Applied Forestry*, 22(3):188–196.

Giglio, L., Descloitres, J., Justice, C. O., and Kaufman, Y. J. (2003). An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment*, 87(2):273–282.

Giglio, L., Schroeder, W., and Justice, C. O. (2016). The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sensing of Environment*, 178:31–41.

Hantson, S., Arneth, A., Harrison, S. P., Kelley, D. I., Prentice, I. C., Rabin, S. S., Archibald, S., Mouillot, F., Arnold, S. R., Artaxo, P., et al. (2016). The status and challenge of global fire modelling. *Biogeosciences*, 13(11):3359–3375.

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A. C. (2007). Long memory in stochastic volatility. In *Forecasting volatility in the financial markets*, pages 351–363. Elsevier.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, volume 52. Cambridge University Press.

Houtman, R. M., Montgomery, C. A., Gagnon, A. R., Calkin, D. E., Dietterich, T. G., McGregor, S., and Crowley, M. (2013). Allowing a wildfire to burn: estimating the effect on future fire suppression costs. *International Journal of Wildland Fire*, 22(7):871–882.

Jin, Y., Goulden, M. L., Faivre, N., Veraverbeke, S., Sun, F., Hall, A., Hand, M. S., Hook, S., and Randerson, J. T. (2015). Identification of two distinct fire regimes in Southern California: Implications for economic impact and future change. *Environmental Research Letters*, 10(9):094005.

Kidwell, K. B. (1991). *NOAA Polar Orbiter Data (TIROS-N, NOAA-6, NOAA-7, NOAA-8, NOAA-9, NOAA-10, NOAA-11, and NOAA-12) Users Guide*. National Oceanic and Atmospheric Administration, National Environmental .

Koopman, S. J., Lit, R., Lucas, A., and Opschoor, A. (2018). Dynamic discrete copula models for high-frequency stock price changes. *Journal of Applied Econometrics*, 33(7):966–985.

Łatuszyński, K., Roberts, G. O., Rosenthal, J. S., et al. (2013). Adaptive Gibbs samplers and related MCMC methods. *The Annals of Applied Probability*, 23(1):66–98.

Li, F., Zhang, X., Kondragunta, S., and Csiszar, I. (2018). Comparison of fire radiative power estimates from viirs and modis observations. *Journal of Geophysical Research: Atmospheres*, 123(9):4545–4563.

Lucas, A., Schwaab, B., and Zhang, X. (2017). Modeling financial sector joint tail risk in the Euro area. *Journal of Applied Econometrics*, 32(1):171–191.

Pastro, L. A., Dickman, C. R., and Letnic, M. (2011). Burning for biodiversity or burning biodiversity? prescribed burn vs. wildfire impacts on plants, lizards, and mammals. *Ecological Applications*, 21(8):3238–3253.

Ressl, R., Lopez, G., Cruz, I., Colditz, R., Schmidt, M., Ressl, S., and Jiménez, R. (2009). Operational active fire mapping and burnt area identification applicable to mexican nature protection areas using modis and noaa-avhrr direct readout data. *Remote Sensing of Environment*, 113(6):1113–1126.

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

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics*, 34(5):2449 – 2495.

Tripathi, R. C., Gupta, P. L., and Gupta, R. C. (1986). Incomplete moments of modified power series districutions 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.

Wintenberger, O. (2013). Continuous invertibility and stable qml estimation of the egarch(1,1) model. *Scandinavian Journal of Statistics*, 40(4):846–867.

Woo, S. H. L., Liu, J. C., Yue, X., Mickley, L. J., and Bell, M. L. (2020). Air pollution from wildfires and human health vulnerability in alaskan communities under climate change. *Environmental Research Letters*, 15(9):094019.

Yang, K., Kang, Y., Wang, D., Li, H., and Diao, Y. (2019). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued autoregressive processes. *Metrika*, 82(7):863–889.

Zamani, H., Faroughi, P., and Ismail, N. (2016). Bivariate generalized Poisson regression model: Applications on health care data. *Empirical Economics*, 51(4):1607–1621.

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.

3.6 Appendix

3.6.1 Proof of the results in Section 3.2

Proof of Proposition 1. To prove Theorem 1, we need some preliminary results.

Let us define

$$S(k,\nu,\lambda)=\sum_{x=0}^{\infty}\frac{(\nu+x\lambda)^{x+k-1}}{x!}e^{-(\nu+x\lambda)},\qquad k=0,1,2,\ldots$$

It has been shown (Consul and Jain, 1973) that $S(k,\nu,\lambda)$ satisfies the following recursion form

$$S(k,\nu,\lambda)=\sum_{i=0}^{\infty}\lambda^{i}(\nu-\lambda i)S(k-1,\nu+\lambda i,\lambda),\qquad k=0,1,2,\ldots$$

from which it is clear that $\nu S(0,\nu,\lambda)=1$, $S(1,\nu,\lambda)=1/(1-\lambda)$ and then also the following recursion holds $S(k,\nu,\lambda)=\nu S(k-1,\nu,\lambda)+\lambda S(k,\nu+\lambda,\lambda)$.

Moreover we define,

$$SP(j,k,\nu,\lambda)=\sum_{x=0}^{\infty}x^{j}\frac{(\nu+x\lambda)^{x+k-1}}{x!}e^{-(\nu+x\lambda)},\qquad k=0,1,2,\ldots$$

From this we have that

$$\begin{aligned}
SP(1,k,\nu,\lambda)&=S(k+1,(\nu+\lambda),\lambda)\\
SP(2,k,\nu,\lambda)&=S(k+2,(\nu+2\lambda),\lambda)+S(k+1,(\nu+\lambda),\lambda)\\
SP(3,k,\nu,\lambda)&=S(k+3,(\nu+3\lambda),\lambda)+3S(k+2,(\nu+2\lambda),\lambda)+S(k+1,(\nu+\lambda),\lambda)
\end{aligned}$$

Since log-likelihood function for a single observation is

$$\log f(y_t)=\log\mu-\log\varphi+(y_t-1)\log\left(\frac{\mu}{\varphi}+y_t\frac{(\varphi-1)}{\varphi}\right)-\log(y_t!)-\frac{\mu}{\varphi}-y_t\frac{(\varphi-1)}{\varphi},$$

we evaluate the first derivative with respect to the parameters μ and φ as

$$\begin{aligned}
\frac{\partial\log f_t}{\partial\mu}&=\frac{\varphi-\mu}{\mu\varphi}+\frac{y_t-1}{\mu+y_t(\varphi-1)}\\
\frac{\partial\log f_t}{\partial\varphi}&=\frac{1}{\varphi}\left[\left(\frac{(y_t-1)\varphi}{\mu+y_t(\varphi-1)}-1\right)\frac{(y_t-\mu)}{\varphi}-1\right]
\end{aligned}$$

Then, given a random variable X distributed as a GP with parameters μ and φ, the expectations of its non-standardized scores with respect to μ and φ can be derived as

$$E\left[\frac{\partial\log f}{\partial\mu}\right]=\frac{\varphi-\mu}{\mu\varphi}+\frac{1}{\varphi}\sum_{x=0}^{\infty}\frac{(x-1)}{\mu/\varphi+x(\varphi-1)/\varphi}f(x;\mu,\varphi)$$

$$= \frac{1}{\varphi}\left[\frac{\varphi-\mu}{\mu} + \frac{\mu}{\varphi}\sum_{x=0}^{\infty} x\frac{(\mu/\varphi + x(\varphi-1)/\varphi)^{x-2}}{x!}e^{-\frac{\mu+x(\varphi-1)}{\varphi}} - \frac{\mu}{\varphi}\sum_{x=0}^{\infty}\frac{(\mu/\varphi + x(\varphi-1)/\varphi)^{x-2}}{x!}e^{-\frac{\mu+x(\varphi-1)}{\varphi}}\right]$$
$$= \frac{1}{\varphi}\left(\frac{\varphi-\mu}{\mu} + \frac{\mu}{\varphi}SP\left(1,-1,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \frac{\mu}{\varphi}S\left(-1,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right)$$
$$= \frac{1}{\varphi}\left[\frac{\varphi-\mu}{\mu} + \frac{\mu}{\varphi}S\left(0,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - S\left(0,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + \frac{(\varphi-1)}{\varphi}S\left(0,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right]$$
$$= \frac{1}{\varphi}\left[\frac{\varphi-\mu}{\mu} + \frac{\mu}{\mu+\varphi-1} - \frac{\varphi}{\mu} + \frac{(\varphi-1)}{\mu+\varphi-1}\right] = 0$$

$$E\left[\frac{\partial \log f}{\partial \varphi}\right] = \frac{1}{\varphi}\left(E\left[\frac{(x-1)(x-\mu)}{\mu+x(\varphi-1)}\right] - \frac{1}{\varphi}(E[x]-\mu) - 1\right)$$
$$= \frac{1}{\varphi}\sum_{x=0}^{\infty}\frac{x^2 - x(\mu+1)+\mu}{\mu+x(\varphi-1)}f(x;\mu,\varphi) - \frac{1}{\varphi}$$
$$= \frac{1}{\varphi^2}\left[\frac{\mu}{\varphi}\sum_{x=0}^{\infty}x^2\frac{(\mu/\varphi + x(\varphi-1)/\varphi)^{x-2}}{x!}e^{-\frac{\mu+x(\varphi-1)}{\varphi}} - \frac{\mu(\mu+1)}{\varphi}\sum_{x=0}^{\infty}x\frac{(\mu/\varphi + x(\varphi-1)/\varphi)^{x-2}}{x!}e^{-\frac{\mu+x(\varphi-1)}{\varphi}} + \right.$$
$$\left. + \frac{\mu^2}{\varphi}\sum_{x=0}^{\infty}x^2\frac{(\mu/\varphi + x(\varphi-1)/\varphi)^{x-2}}{x!}e^{-\frac{\mu+x(\varphi-1)}{\varphi}} - \varphi\right]$$
$$= \frac{1}{\varphi^2}\left[\frac{\mu}{\varphi}SP\left(2,-1,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \frac{\mu(\mu+1)}{\varphi}SP\left(1,-1,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + \frac{\mu^2}{\varphi}S\left(-1,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \varphi\right]$$
$$= \frac{1}{\varphi^2}\left[\frac{\mu}{\varphi}S\left(1,\frac{\mu+2(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + \frac{\mu}{\varphi}S\left(0,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \right.$$
$$- \frac{\mu(\mu+1)}{\varphi}S\left(0,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + \mu S\left(0,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) -$$
$$\left. - \frac{\mu(\varphi-1)}{\varphi}S\left(0,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \varphi\right]$$
$$= \frac{1}{\varphi^2}\left[\mu + \frac{\mu}{\mu+\varphi-1} - \frac{\mu(\mu+1)}{\mu+\varphi-1} + \varphi - \frac{\mu(\varphi-1)}{\mu+\varphi-1} - \varphi\right] = 0$$

The information matrix of the parameters of the distribution can be derived as the expectation of the negative second derivative of the likelihood with respect to the parameters. The second derivative of the likelihood with respect to μ and φ are

i. $\frac{\partial^2 \log f_t}{\partial \mu^2} = -\frac{1}{\mu^2} - \frac{(y_t-1)}{(\mu+y_t(\varphi-1))^2}$

ii. $\frac{\partial^2 \log f_t}{\partial \varphi^2} = \frac{1}{\varphi^2}\left[\frac{2(y_t-\mu)}{\varphi} - \frac{(y_t-1)(y_t-\mu)(\mu+y_t(2\varphi-1))}{(\mu+y_t(\varphi-1))^2} + 1\right]$

iii. $\frac{\partial^2 \log f_t}{\partial \varphi \partial \mu} = -\frac{1}{\varphi^2} - \frac{x(x-1)}{(\mu+x(\varphi-1))^2}$

Taking the expectation with respect to x one obtains:

i. $E\left[\frac{\partial^2 \log f_t}{\partial \mu^2}\right] = -\frac{1}{\mu^2} - E\left[\frac{(x-1)}{(\mu+x(\varphi-1))^2}\right]$
$$= -\frac{1}{\mu^2} - \sum_{x=0}^{\infty}\frac{(x-1)}{(\mu+x(\varphi-1))^2}f(x;\mu,\varphi)$$
$$= -\frac{1}{\mu^2} - \frac{\mu}{\varphi^3}\sum_{x=0}^{\infty}(x-1)\frac{(\mu/\varphi + x(\varphi-1)/\varphi)^{x-3}}{x!}e^{-\frac{\mu+x(\varphi-1)}{\varphi}}$$

$$
\begin{aligned}
&= -\frac{1}{\mu^2} - \frac{\mu}{\varphi^3}\left[SP\left(1,-2,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - S\left(-2,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right] \\
&= -\frac{1}{\mu^2} - \frac{\mu}{\varphi^3}\left[\frac{\mu+\varphi-1}{\mu}S\left(-1,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \frac{\varphi}{\mu}S\left(-1,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right] \\
&= -\frac{1}{\mu^2} - \frac{\mu+\varphi-1}{\varphi^3}\left[\frac{\varphi}{\mu+\varphi-1}S\left(0,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \frac{(\varphi-1)}{\mu+\varphi-1}S\left(0,\frac{\mu+2(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right] + \\
&+ \frac{1}{\varphi^2}\left[\frac{\varphi}{\mu}S\left(0,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \frac{(\varphi-1)}{\mu}S\left(0,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right] \\
&= -\frac{1}{\mu^2} + \frac{(\varphi-1)}{\varphi^3}S\left(0,\frac{\mu+2(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \frac{\mu+\varphi-1}{\mu\varphi^2}S\left(0,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + \\
&+ \frac{1}{\mu\varphi}S\left(0,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) \\
&= -\frac{1}{\mu^2} + \frac{(\varphi-1)}{\varphi^2(\mu+2(\varphi-1))} - \frac{1}{\mu\varphi} + \frac{1}{\mu^2} \\
&= -\frac{\mu+2\varphi(\varphi-1)}{\mu\varphi^2(\mu+2(\varphi-1))}
\end{aligned}
$$

ii. $$
\begin{aligned}
E\left[\frac{\partial^2 \log f_t}{\partial\varphi^2}\right] &= \frac{1}{\varphi^2}\left(1 - \frac{2\mu}{\varphi} + \frac{2}{\varphi}E[x] + [2\mu(\varphi-1)+(2\varphi-1)]E\left[\frac{x^2}{(\mu+2(\varphi-1))^2}\right] - \right. \\
&- (2\varphi-1)E\left[\frac{x^3}{(\mu+2(\varphi-1))^2}\right] - [2\mu(\varphi-1)-\mu^2]E\left[\frac{x}{(\mu+2(\varphi-1))^2}\right] - \\
&\left. -\mu^2 E\left[\frac{1}{(\mu+2(\varphi-1))^2}\right]\right) \\
&= \frac{1}{\varphi^2} + \frac{\mu}{\varphi^5}\left\{[2\mu(\varphi-1)+(2\varphi-1)]SP\left(2,-2,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - (2\varphi-1)SP\left(3,-2,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \right. \\
&\left. - [2\mu(\varphi-1)-\mu^2]SP\left(1,-2,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \mu^2 SP\left(0,-2,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right\} \\
&= \frac{1}{\varphi^2} + \frac{\mu}{\varphi^5}\left\{[2\mu(\varphi-1)+(2\varphi-1)]\left[S\left(0,\frac{\mu+2(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + S\left(-1,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right] - \right. \\
&- (2\varphi-1)\left[S\left(1,\frac{\mu+3(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + 3S\left(0,\frac{\mu+2(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + \right. \\
&\left. \left. + S\left(-1,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right] - [2\mu(\varphi-1)-\mu^2]S\left(-1,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \right. \\
&\left. -\mu^2 S\left(-2,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right\} \\
&= \frac{1}{\varphi^2} + \frac{\mu}{\varphi^5}\left\{[2\mu(\varphi-1)+(2\varphi-1)]S\left(0,\frac{\mu+2(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \right. \\
&- (2\varphi-1)S\left(1,\frac{\mu+3(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + \mu^2 S\left(-1,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \\
&\left. -\mu^2\left[\frac{\varphi}{\mu}S\left(-1,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \frac{(\varphi-1)}{\mu}S\left(-1,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right]\right\} \\
&= \frac{1}{\varphi^2} + \frac{\mu}{\varphi^5}\left\{2\varphi\frac{2\mu(\varphi-1)+(2\varphi-1)}{\mu+2(\varphi-1)} - \varphi(2\varphi-1) - \varphi^2 S\left(0,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + \right. \\
&\left. + [\mu\varphi+\varphi(\varphi-1)]S\left(0,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \mu(\varphi-1)S\left(0,\frac{\mu+2(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right\} \\
&= \frac{\mu}{\varphi^4}[\varphi-(2\varphi-1)] + \frac{\mu}{\varphi^4(\mu+2(\varphi-1))}[\mu(\varphi-1)-2(2\varphi-1)] \\
&= \frac{\mu}{\varphi^4}\left[\frac{\mu(\varphi-1)-2(2\varphi-1)}{\mu+2(\varphi-1)} - (\varphi-1)\right] \\
&= -\frac{2\mu}{\varphi^2(\mu+2(\varphi-1))}
\end{aligned}
$$

iii. $$E\left[\frac{\partial^2 \log f_t}{\partial\varphi\partial\mu}\right] = -\frac{1}{\varphi^2} - E\left[\frac{x\,(x-1)}{(\mu + x\,(\varphi-1))^2}\right]$$
$$= -\frac{1}{\varphi^2} - \frac{\mu}{\varphi^3}\left[SP\left(2,-2,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - SP\left(1,-2,\frac{\mu}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right]$$
$$= -\frac{1}{\varphi^2} - \frac{\mu}{\varphi^3}\left[S\left(0,\frac{\mu+2\,(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right) + S\left((-1,\frac{\mu+\varphi-1}{\varphi},\frac{(\varphi-1)}{\varphi}\right) - \right.$$
$$\left. -S\left(-1,\frac{\mu+2\,(\varphi-1)}{\varphi},\frac{(\varphi-1)}{\varphi}\right)\right]$$
$$= -\frac{1}{\varphi^2} - \frac{\mu\varphi}{\varphi^3\,(\mu+2\,(\varphi-1))}$$
$$= -\frac{2\,(\varphi-1)}{\varphi^2\,(\mu+2\,(\varphi-1))}.$$

The information matrix can be derived as

$$\begin{aligned}
\mathcal{I}_{\mu\mu} &= -E\left[\frac{\partial^2 \log f_t}{\partial\mu^2}\right] = \frac{\mu + 2\varphi\,(\varphi-1)}{\mu\varphi^2\,(\mu+2\,(\varphi-1))},\\
\mathcal{I}_{\varphi\varphi} &= -E\left[\frac{\partial^2 \log f_t}{\partial\varphi^2}\right] = \frac{2\mu}{\varphi^2\,(\mu+2\,(\varphi-1))},\\
\mathcal{I}_{\mu\varphi} &= -E\left[\frac{\partial^2 \log f_t}{\partial\varphi\partial\mu}\right] = \frac{2\,(\varphi-1)}{\varphi^2\,(\mu+2\,(\varphi-1))}.
\end{aligned}$$

□

Proof of Proposition 2. Keeping in mind the two link functions $\mu = e^{\theta_1}$ and $\varphi = 1 + e^{\theta_2}$, to evaluate the information matrix for the parameters θ_1 and θ_2 we start with the derivatives:

$$\begin{aligned}
\frac{\partial \log f_t}{\partial\theta_1} &= \frac{\partial \log f_t}{\partial\mu}\frac{\partial\mu}{\partial\theta_1} = \frac{\partial \log f_t}{\partial\mu}\mu,\\
\frac{\partial \log f_t}{\partial\theta_2} &= \frac{\partial \log f_t}{\partial\varphi}\frac{\partial\varphi}{\partial\theta_2} = \frac{\partial \log f_t}{\partial\varphi}\,(\varphi-1)\,.
\end{aligned}$$

The second derivatives of the likelihood with respect to θ_1 and θ_2 are

$$\begin{aligned}
\frac{\partial^2 \log f_t}{\partial\theta_1^2} &= \frac{\partial \log f_t}{\partial\mu}\frac{\partial\mu}{\partial\theta_1} + \frac{\partial^2 \log f_t}{\partial\mu}\mu^2,\\
\frac{\partial^2 \log f_t}{\partial\theta_2^2} &= \frac{\partial \log f_t}{\partial\varphi}\frac{\partial\varphi}{\partial\theta_2} + \frac{\partial^2 \log f_t}{\partial\varphi}\,(\varphi-1)^2\,.
\end{aligned}$$

The information matrix can be derived as

$$\begin{aligned}
\mathcal{I}_{\theta_1\theta_1} &= \mathcal{I}_{\mu\mu}\mu^2 = \frac{\mu\,(\mu+2\varphi\,(\varphi-1))}{\varphi^2\,(\mu+2\,(\varphi-1))},\\
\mathcal{I}_{\theta_2\theta_2} &= \mathcal{I}_{\varphi\varphi}\,(\varphi-1)^2 = \frac{2\mu\,(\varphi-1)^2}{\varphi^2\,(\mu+2\,(\varphi-1))}.
\end{aligned}$$

□

Proof of Proposition 3. To prove Proposition 3 we need a preliminary result about log moments. Given a strictly positive sequence $\{a_n\}_{n\in\mathbb{N}}$, where its series its convergent as

$$\sum_{n=0}^{\infty} n\, a_n = L \leq \infty$$

we have that

$$\log^+ n\, a_n \leq n\, a_n,$$
$$\sum_{n=0}^{\infty} \log^+ n\, a_n \leq \sum_{n=0}^{\infty} n\, a_n.$$

Then if

$$a_n = \frac{\mu}{\varphi} \frac{\left(\mu/\varphi + n\left(\varphi-1\right)/\varphi\right)^{n-1}}{n!} e^{-\frac{\mu+n(\varphi-1)}{\varphi}}$$

we have that

$$E\left[\log^+ y_t\right] \leq E\left[y_t\right] \leq \infty.$$

Let's consider a compact parameter space of the form

$$\Psi = [\omega_1^-, \omega_1^+] \cdot [\phi_1^-, \phi_1^+] \cdot [\kappa_1^-, \kappa_1^+] \cdot [\varphi^-, \varphi^+]$$

Hence, the standardised score is

$$u_{1,t} = \left[\frac{\mu_{t|t-1}(\theta_{1,t|t-1}) - \varphi}{\mu_{t|t-1}(\theta_{1,t|t-1})\varphi} - \frac{(y_t - 1)}{\mu_{t|t-1}(\theta_{1,t|t-1}) + y_t(\varphi-1)}\right] \frac{\varphi^2(\mu_{t|t-1}(\theta_{1,t|t-1}) + 2(\varphi-1))}{\mu_{t|t-1}(\theta_{1,t|t-1}) + 2\varphi(\varphi-1)}$$

where $\mu_{t|t-1}(\theta_{1,t|t-1}) = \exp(\theta_{1,t|t-1})$, after suppressing the time subscript becomes

$$u_{1,t} = \frac{\varphi[\mu(\theta_1)^2 - y(\mu(\theta_1) + \varphi(\varphi-1))](\mu(\theta_1) + 2(\varphi-1))}{(\mu(\theta_1) + y(\varphi-1))(\mu(\theta_1) + 2\varphi(\varphi-1))}$$

and x_t is defined as

$$x_t\left(\theta_{1,t|t-1}, \boldsymbol{\psi}\right) = \frac{\partial g\left(\theta_{1,t|t-1}, y_t, \boldsymbol{\psi}\right)}{\partial \theta_{1,t|t-1}} = \phi_1 + \kappa_1 \frac{\partial u_{1,t}}{\partial \theta_{1,t|t-1}}$$

where suppressing the time subscripts

$$\begin{aligned}\frac{\partial u_1}{\partial \theta_1} =& \frac{\varphi^2(\mu(\theta_1) + 2(\varphi-1))}{(\mu(\theta_1) + 2\varphi(\varphi-1))}\left[\frac{y(y-1)(\varphi-1)}{(\mu(\theta_1) + y(\varphi-1))^2} - \frac{1}{\varphi} - \right.\\ &\left. - \left(\frac{\varphi - \mu(\theta_1)}{\varphi} + \frac{\mu(\theta_1)(y-1)}{\mu(\theta_1) + y(\varphi-1)}\right)\left(\frac{1}{\mu(\theta_1) + 2\varphi(\varphi-1)} - \frac{2(\varphi-1)}{\mu(\theta_1)(\mu(\theta_1) + 2(\varphi-1))}\right)\right].\end{aligned}$$

To verify condition i) we make use of Lemma 2.2 of Straumann and Mikosch (2006)

$$E\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|g\left(\hat{\theta}_{1,1},y_t,\boldsymbol{\psi}\right)\right|\right)\leq 4\log 2+E\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}|(1-\phi_1)\omega_1|\right)+E\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}|\phi_1|\right)$$
$$+E\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\hat{\theta}_{1,1}\right|\right)++E\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}|\kappa_1|\right)+E\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}|u_{1,1}|\right)<\infty$$

which is satisfied taking the supremum of $\boldsymbol{\psi}$ in the compact space Ψ and for bounded $E\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}|u_{1,1}|\right)$. Since

$$E\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}|u_{1,t}|\right)\leq E\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\frac{\varphi[\hat{\mu}(\hat{\theta}_1)^2-y(\hat{\mu}(\hat{\theta}_1)+\varphi(\varphi-1))](\hat{\mu}(\hat{\theta}_1)+2(\varphi-1))}{(\hat{\mu}(\hat{\theta}_1)+y(\varphi-1))(\hat{\mu}(\hat{\theta}_1)+2\varphi(\varphi-1))}\right|\right)$$

Applying Lemma 2.2 of Straumann and Mikosch (2006) we have that

$$\leq 2\log 2+\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\hat{\mu}(\hat{\theta}_1)^2\right|\right)+E\left[\log^+|y|\right]+\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\hat{\mu}(\hat{\theta}_1)+\varphi(\varphi-1)\right|\right)+$$
$$+\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\varphi(\hat{\mu}(\hat{\theta}_1)+2(\varphi-1))\right|\right)-2\log 2-\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\hat{\mu}(\hat{\theta}_1)\right|\right)-E\left[\log^+|y|\right]-$$
$$-E\log^+|(\varphi-1)|-\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\hat{\mu}(\hat{\theta}_1)+2\varphi(\varphi-1)\right|\right)$$
$$\leq\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\hat{\mu}(\hat{\theta}_1)^2\right|\right)+\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\hat{\mu}(\hat{\theta}_1)+\varphi(\varphi-1)\right|\right)+\log^+\left(\sup_{\boldsymbol{\psi}\in\Psi}\left|\varphi(\hat{\mu}(\hat{\theta}_1)+2(\varphi-1))\right|\right)$$

which is satisfied taking the supremum of $\boldsymbol{\psi}$ in the compact space Ψ and for the bounded logarithmic moment $E\log^+|y|<\infty$. Condition ii) can be also verified by the use of the Lemma 2.2 of Straumann and Mikosch (2006) to have that

$$E\sup_{\boldsymbol{\psi}\in\Psi}\sup_{\theta_1\in\mathcal{F}_\psi}\log^+|x_t(\theta_1,\boldsymbol{\psi})|\leq 2\log 2+\sup_{\boldsymbol{\psi}\in\Psi}\log^+|\phi_1|+\sup_{\boldsymbol{\psi}\in\Psi}\log^+|\kappa_1|+$$
$$+E\sup_{\boldsymbol{\psi}\in\Psi}\sup_{\theta_1\in\mathcal{F}_\psi}\log^+\left|\frac{\partial u_1}{\partial\theta_1}\right|<\infty$$

which is satisfied by taking the supremum of $\boldsymbol{\psi}$ in the compact space Ψ and for bounded $E\sup_{\boldsymbol{\psi}\in\Psi}\sup_{\theta_1\in\mathcal{F}_\psi}\log^+\left|\frac{\partial u_1}{\partial\theta_1}\right|$. Using Lemma 2.2 of Straumann and Mikosch (2006) we have that

$$E\sup_{\boldsymbol{\psi}\in\Psi}\sup_{\theta_1\in\mathcal{F}_\psi}\log^+\left|\frac{\partial u_1}{\partial\theta_1}\right|\leq E\sup_{\boldsymbol{\psi}\in\Psi}\sup_{\theta_1\in\mathcal{F}_\psi}\log^+\left|\frac{\varphi^2(\mu_{t|t-1}(\theta_1)+2(\varphi-1))}{(\mu_{t|t-1}(\theta_1)+2\varphi(\varphi-1))}\right|+$$
$$+E\sup_{\boldsymbol{\psi}\in\Psi}\sup_{\theta_1\in\mathcal{F}_\psi}\log^+\left|\frac{y(y-1)(\varphi-1)}{(\mu(\theta_1)+y(\varphi-1))^2}-\frac{1}{\varphi}+\right.$$
$$+\frac{\mu(\theta_1)-\varphi}{\varphi}\left(\frac{1}{\mu(\theta_1)+2\varphi(\varphi-1)}-\frac{2(\varphi-1)}{\mu(\theta_1)(\mu(\theta_1)+2(\varphi-1))}\right)-$$

$$
\begin{aligned}
&-\frac{\mu(\theta_1)(y-1)}{\mu(\theta_1)+y(\varphi-1)}\left(\frac{1}{\mu(\theta_1)+2\varphi(\varphi-1)}-\frac{2(\varphi-1)}{\mu(\theta_1)(\mu(\theta_1)+2(\varphi-1))}\right)\Bigg| \\
&\leq E\sup_{\boldsymbol{\psi}\in\Psi}\sup_{\theta_1\in\mathcal{F}_\psi}\log^+\left|\frac{\varphi^2(\mu_{t|t-1}(\theta_1)+2(\varphi-1))}{(\mu_{t|t-1}(\theta_1)+2\varphi(\varphi-1))}\right|+ \\
&+E\sup_{\boldsymbol{\psi}\in\Psi}\sup_{\theta_1\in\mathcal{F}_\psi}\log^+\left|\frac{y(y-1)(\varphi-1)}{(\mu(\theta_1)+y(\varphi-1))^2}+\frac{\mu(\theta_1)}{\varphi(\mu(\theta_1)+2\varphi(\varphi-1))}+\right. \\
&+\frac{2\varphi(\varphi-1)}{\mu(\theta_1)(\mu(\theta_1)+2(\varphi-1))}+\frac{2(y-1)(\varphi-1)}{(\mu(\theta_1)+y(\varphi-1))(\mu(\theta_1)+2(\varphi-1))}- \\
&\left.-\frac{1}{\varphi}\right|
\end{aligned}
$$

and taking the supremum of θ_1 in the space $\mathcal{F}_\psi$ we have that

$$
\leq \sup_{\varphi\in\Psi}\log^+\left(\varphi^2\right)+\sup_{\varphi\in\Psi}\log^+\left(\frac{1}{\varphi}-\frac{1}{\varphi}\right)<\sup_{\varphi\in\Psi}\log^+\left(\varphi^2\right)<\infty
$$

which is satisfied taking the supremum of $\boldsymbol{\psi}$ in the compact space Ψ. Condition *iii*) can be also verified by the use of Lemma 2.2 of Straumann and Mikosch (2006) to have that

$$
\begin{aligned}
E\log\Lambda_t(\boldsymbol{\psi}) &\leq E\log^+\sup_{\theta_1\in\mathcal{F}_\psi}\left|x_t\left(\theta_1,\boldsymbol{\psi}\right)\right| \\
&\leq E\sup_{\theta_1\in\mathcal{F}_\psi}\log^+\left|\phi_1+\kappa_1\frac{\partial u_1}{\partial\theta_1}\right| \\
&\leq \log^+|\phi_1|\leq 0
\end{aligned}
$$

that implies $|\phi_1|<1$, which, following Harvey (2013), implies also the stationarity of θ_1 completing the proof.

□

Chapter 4

Generalised Lagrangian Katz Integer-Valued Autoregressive Process [1]

Abstract. A new integer-valued autoregressive process (INAR) with Generalised Lagrangian Katz (GLK) innovations is defined. We show that our GLK-INAR process is stationary, discrete semi-self-decomposable, infinite divisible, and provides a flexible modeling framework for count data allowing for under- and over-dispersion, asymmetry, and excess of kurtosis. A Bayesian inference framework and an efficient posterior approximation procedure based on Markov Chain Monte Carlo are provided. The proposed model family is applied to a Google Trend dataset which proxies the public concern about climate change around the world. The empirical results show that the larger flexibility of the GLK-INAR returns a better fitting in comparison with a simple INAR model on more than half of the 130 time series considered. The application provides new evidence of heterogeneity across countries and keywords in the persistence, uncertainty, and long-run public awareness level about climate change.

Keywords: Bayesian inference, Big Data, Counts time series, Climate Risk, Generalized Lagrangian Katz distribution.

4.1 Introduction

In the recent years there has been a large interest in discrete-time integer-valued models, also due to increased availability of count data in very diverse fields including finance (Liesenfeld et al., 2006; Rydberg and Shephard, 2003; Aknouche et al., 2021), economics (Freeland, 1998; Freeland and McCabe, 2004; Berry and West, 2020), social sciences (Pedeli and Karlis, 2011), sports (Shahtahmassebi and Moyeed, 2016), image processing (Afrifa-Yamoah and Mueller, 2022) and

[1] In collaboration with Federico Bassetti (Polytechnic University of Milan, Italy).

oceanography (Cunha et al., 2018). Among the modelling approaches, integer-valued autoregressive processes (INAR), introduced independently by Al-Osh and Alzaid (1987) and McKenzie (1985), become vary popular. The stochastic construction of the INAR relies on the binomial thinning operator and the properties of the model on the discrete self-decomposability of the stationary distribution of the process (Steutel and van Harn, 1979). See also Weiß (2008) and Scotto et al. (2015) for a review on thinning operators.

The original INAR model has been studied further in Al-Osh and Alzaid (1987) and has been extend along different directions. (McKenzie, 1986) introduced an INAR model with negative-binomial and geometric marginal distributions, Jin-Guan and Yuan (1991) extended the INAR(1) model of Al-Osh and Alzaid (1987) to the higher order INAR(p). Al-Osh and Aly (1992) introduced a negative-binomial INAR with a new iterated thinning operator. Other extensions of the INAR process has been made in order to include a seasonal structure in the model (e.g., see Bourguignon et al., 2016). INAR models with valued in the set of signed integers have been propose firstly by Kim and Park (2008) and generalised by Alzaid and Omair (2014) and Andersson and Karlis (2014). Freeland (2010) proposed a true integer-valued autoregressive model (TINAR(1)). More flexible INAR models have been introduced by assuming more flexible distributions for the innovations terms. Alzaid and Al-Osh (1993) propose integer-valued ARMA process with Generalized Poisson marginals and Kim and Lee (2017) introduced INAR with Katz innovations.

In this chapter we extend further the Negative Binomial, Generalized Poisson and Katz INAR processes by assuming a more general distribution for the innovations which includes some well-known distributions used in time series of count data and some distributions which have not yet been used in modelling time series of counts. The two-parameter Katz distribution belongs to the Lagrangian Katz family together with the three-parameter Lagrangian Katz (e.g., see Consul and Famoye, 2006, ch. 12). The Lagrangian Katz is a flexible distribution and naturally arises as first crossing probabilities, which is a common problem in actuarial mathematics, e.g. claim number distribution in cascading processes or ruin probability in discrete-time risk models. The Lagrangian Katz distribution has been extended further by Janardan (1998) and Janardan (1999) which introduced the four-parameter generalized Pólya-Eggenberger (GPED) distributions of the first and second kind. Janardan (1998) showed that both families contain the Lagrangian Katz distribution as a special case. See also Johnson et al. (2005), ch. 2,5 and 7, for a review on the relationships between Lagrangian Katz distributions and other discrete distribution families. In this chapter we consider the four-parameter GPED of the first kind, or Generalized Lagrangian Katz (GLK) distribution in the following, since it enjoys the properties of the Lagrangian distributions and includes some well-known distributions such as Generalized Poisson and Negative Binomial as special cases.

Another contribution of this chapter regards the inference method. Various approaches to inference have traditionally presented for count data models, such

as conditional likelihood approach, generalized method of moments and Yule-Walker approach. See Weiß and Kim (2013) for a review. Despite the popularity gained in the recent years by Bayesian methods, the applications to count data models are still limited (e.g., see McCabe and Martin, 2005; Neal and Subba Rao, 2007; Drovandi et al., 2016; Shang and Zhang, 2018; Garay et al., 2020b). Thus, we provide a Bayesian inference procedure for our model and illustrate the efficiency of the procedure on a synthetic dataset. The Bayesian approach to inference entirely considers parameter uncertainty in the prior knowledge about a random process and allows for imposing parameter restrictions through the specification of the prior distribution (Chen and Lee, 2016). The posterior distribution of the parameters quantifies uncertainty in the estimation (Chen and Lee, 2017), which can be included in the prediction. The inference from the Bayesian perspective may result in richer inferences in the case of small samples (Garay et al., 2020a) and extra-sample information and in robust inference in the presence of outliers (Fried et al., 2015). Finally, model selection for both nested and non-nested models can be easily carried out.

We illustrate the model's flexibility with an application to an original Google Trend dataset on measures of public concern about climate change in different countries. Assessing the level of public awareness and knowledge in a specific topic and understanding the dynamics in the social consciousness allows for designing more effective public policies. For this reason, researchers measured and studied the level of awareness about the effects of climate change in different sectors of society such as households (Frondel et al., 2017), winegrowers (Battaglini et al., 2009), farmers (Fahad and Wang, 2018; Singh et al., 2017), mountain peoples (Ullah et al., 2018). Most of these studies rely on surveys conducted in a specific geographical area and sector of society, with a few exceptions. For example, Ziegler (2017) proposed a cross-country analysis of climate change beliefs and attitudes. Lineman et al. (2015) provided a broader and global perspective by exploiting the potentiality of big data provided by Google Trend. This extended climate change perception literature along two lines. First, we consider a multi-country dataset including country-specific measures to capture heterogeneity across the world in public awareness. Moreover, we offer a model-based approach and an inference procedure to analyze these measures, which allow for a better comprehension of the cross-country and cross-keyword heterogeneity in the long-run level of public perception and perception uncertainty and persistence.

The chapter is organized as follows. Section 2 introduces the Generalized Lagrangian Katz family, the INAR process, and some of their properties. Section 3 proposes a Bayesian inference procedure and provides some simulation results. Section 4 illustrates the model and inference framework on a multi-country Google Trend dataset on climate change and provides some new findings in the public interest in climate change. Section 5 concludes.

4.2 INAR(1) with generalized Katz innovations

4.2.1 Generalized Lagrangian Katz family

The Generalized Lagrangian Katz (GLK) distribution has been introduced by (Janardan, 1998), which named the distribution Generalized Polya Eggenberger distribution. (Consul and Famoye, 2006) argued that since the distribution is not related to the Polya, it should be named Generalized Lagrangian Katz distribution. The GLK has probability mass function (pmf), $P(X = x) = p_x$, defined as

$$p_x = \frac{1}{x!}\beta^x \frac{a}{c} \frac{1}{(\frac{a}{c} + x\frac{b}{c} + x)} (1-\beta)^{\frac{a}{c}+x\frac{b}{c}} \left(\frac{a}{c} + x\frac{b}{c} + 1\right)_{x\uparrow} \quad x = 0, 1, 2, \ldots \tag{4.1}$$

where $(x)_{k\uparrow} = x(x+1)\ldots(x+k-1)$ is the rising factorial with the convention that $(x)_0 = 1$, and $a > 0$, $c > 0$, $b \geq 0$ and $0 < \beta < 1$ are the parameters. In the following, we denote the distribution with $\mathcal{GLK}(a, b, c, \beta)$.

The GLK distribution has probability generating function (pgf)

$$H(u) = \sum_{x=0}^{\infty} p_x u^x \tag{4.2}$$

which satisfies:

$$H(u) = (1 - \beta + \beta z)^{a/c}, \; z = u(1 - \beta + \beta z)^{b/c+1}, \tag{4.3}$$

or alternatively

$$H(u) = \left(\frac{1-\beta}{1-\beta z}\right)^{a/c}, \; z = u\left(\frac{1-\beta z}{1-\beta}\right)^{b/c}, \tag{4.4}$$

see Janardan (1998).

Remark 1 (Lagrangian family)**.** *The definition of GLK given in Janardan (1998) is obtained as a special case of "generalized Lagrangian distribution" starting from the Lagrangian expansion*

$$H(u) = f(0) + \sum_{x=1}^{\infty} \frac{u^x}{x!} \left|\partial^{x-1}(g^x(z)f'(z))\right|_{z=0}$$

where $H(u) = f(z)$ for u satisfying $z = ug(z)$, (Consul and Famoye, 2006, p. 10-11). Choosing $g(z) = ((1-\beta)/(1-\beta z))^{b/\beta}$ and $f(z) = ((1-\beta)/(1-\beta z))^{a/c}$, one gets (4.3) *and* (4.1) *follows after some algebra. See Appendix 4.6.1 for details.*

The following remarks clarify the relationship with other definitions of Katz distributions and with other discrete distributions. For some values of the parameters, the GLK distribution reduces either to some well known distributions in time series analysis, or to distributions which have not yet been used in count data modelling.

Remark 2. *Three distributions related to our research are: (i) the Lagrangian Katz distribution $\mathcal{LK}(a, b, \beta)$ by replacing c with β, (which is called Generalized Katz in (Consul and Famoye, 2006)); (ii) the Katz distribution $\mathcal{K}(a, \beta)$ for $b = 0$ and by replacing c with β (Katz, 1965); (iii) the Polya-Eggenberger distribution $\mathcal{PE}(a, c, \beta)$ for $b = 0$ (Janardan, 1998). Note that the Katz distribution in Consul and Famoye (2006), Tab. 2.1, is not the Katz distribution of Katz (1965), it corresponds instead to the Generalized Polya Eggenberger of the first type ($GPED_1$-I) of Janardan (1998) and can be obtain as the limit of the zero-truncated GLK for $a \to -c$.*

The probability mass function of the GLK for different parameter settings given in Fig. 4.1. In the top-left plot we compare $\mathcal{K}(a, c)$, $\mathcal{LK}(a, b, \beta)$ and $\mathcal{GLK}(a, b, c, \beta)$ with the same mean. The top-right plot illustrates the sensitivity of the $\mathcal{GLK}(a, b, c, \beta)$ pmf with respect to the different parameters. All distributions have the same mean (vertical dashed line). The bottom plots illustrate the effects of the parameters on the tails (log-scale) for a $\mathcal{GLK}(a, b, c, \beta)$ with over-dispersion $VMR = 50/15$ (left) and under-dispersion $VMR = 13/15$ (right).

Remark 3. *The following standard distributions have been used to define INAR processes: (i) the Negative Binomial distribution $\mathcal{NB}(r, p)$ for $b = 0$ $\beta = 1-p$ and $r = a/c$; (ii) the Binomial distribution $\mathcal{B}in(n, p)$ for $c = 1$, $b = -1$, $a = n \in \mathbb{N}$ and $\beta = p$; the Poisson distribution $\mathcal{P}(\theta)$ for $c \to 0$, $b \to 0$ s.t. $a\beta/c = \theta$.*

Remark 4. *The GLK family includes also the generalizations of the distributions given in the previous remark, that are the Generalized Negative Binomial distribution $\mathcal{GNB}(r, \gamma, p)$ for $c = 1$, $a = r$, $b = \gamma - 1$ and $\beta = p$ and the Generalized Poisson (GP) distribution $\mathcal{GP}(\theta, \lambda)$ for $c \to 0$ s.t. $\frac{b}{a} = \lambda$ and $a\beta/c = \theta > 0$ with $0 < \lambda < \theta^{-1}$. The GP limit of the GLK distribution is stated in (Consul and Famoye, 2006) without proof. In Appendix 4.6.1 we provide a proof of the result.*

Studying the moments allows for a better understanding of the flexibility of the GLK distribution. Four moments relevant to our analysis are the following.

Proposition 1. *Let $X \sim \mathcal{GLK}(a, b, c, \beta)$, define $\mu'_k = \mathbb{E}((X - \mathbb{E}(X))^k)$ and $\mu_k = \mathbb{E}(X^k)$ then*

$$\begin{aligned}
\mu_1 &= \frac{a\theta}{\kappa}, \qquad \mu'_2 = \frac{(1-\beta)a\theta}{\kappa^3}, \qquad \mu'_3 = \frac{a\theta(1-2\beta)(1-\beta)}{\kappa^4} + \frac{3a\theta^2(1-\beta)^2(b+c)}{\kappa^5} \\
\mu'_4 &= a\theta(1-\beta)(1+2\theta b-(b+c)\beta\theta)\left(\frac{1-\beta-\beta^2}{\kappa^6} + \frac{5a\theta(1-\beta)(b+c)}{\kappa^7}\right) + 3(\mu'_2)^2
\end{aligned}$$

where $\kappa = 1 - \beta - b\beta/c$ and $\theta = \beta/c$.

For a proof, see Janardan (1998) Theorems 1-3. Another quantity of interest is the coefficient of variation which represents a measure of relative dispersion,

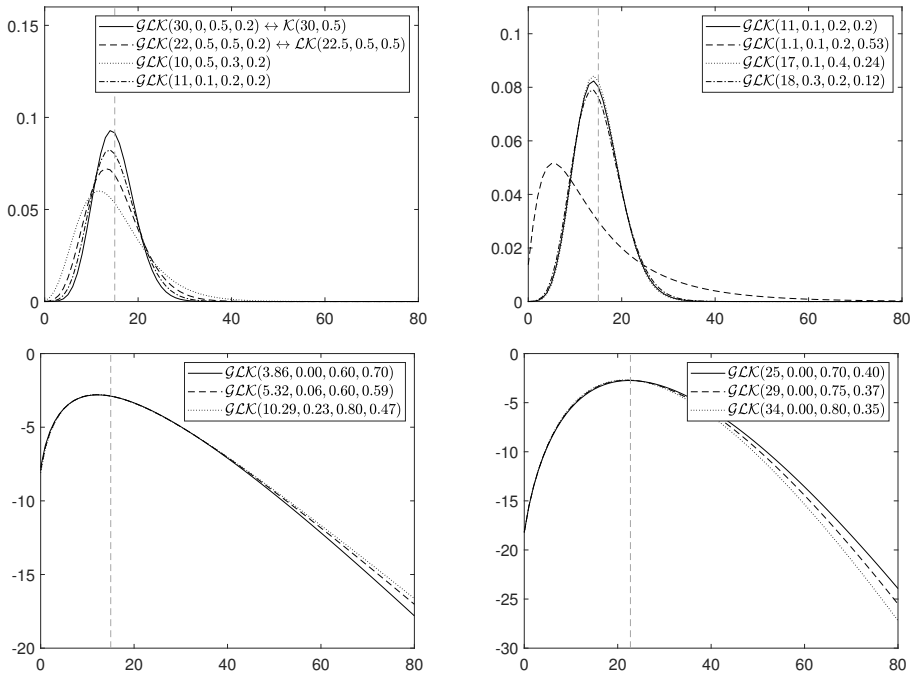


Figure 4.1: Probability mass function of the Generalized Lagrangian Katz for different parameter settings. Top left: comparison between $\mathcal{LK}(a,c)$, $\mathcal{LK}(a,b,\beta)$ and $\mathcal{GLK}(a,b,c,\beta)$. Top right: sensitivity of $\mathcal{GLK}(a,b,c,\beta)$ with respect to the parameters. Bottom: effect of the parameters on the tails (log scale) for a $\mathcal{GLK}(a,b,c,\beta)$ with over-dispersion ($VMR = 50/15$, left) and under-dispersion ($VMR = 13/15$, right). In each plot the distribution mean (vertical dashed line).

given by the standard deviation to mean ratio. From the previous result it follows that the coefficient of variation is

$$CV = \left(\frac{(1-\beta)}{a\theta\kappa}\right)^{1/2}.$$

Another classical measure of dispersion is the Fisher index, given by the variance-to-mean ratio

$$VMR = \frac{(1-\beta)}{\kappa^2},$$

which does not depend on the parameter a. For a given β, following the values of κ (b and c), the distribution allows for various degree of dispersion: not dispersed ($VMR = 0$), under-dispersed ($VMR < 1$), equally dispersed ($VMR = 1$) and over-dispersed ($VMR > 1$).

The skewness and the kurtosis of the distribution are

$$S = \frac{(1-2\beta)\kappa^{1/2}}{((1-\beta)a\theta)^{1/2}} + \frac{3((1-\beta)\theta)^{1/2}(b+c)}{a\kappa^{1/2}}, \tag{4.5}$$

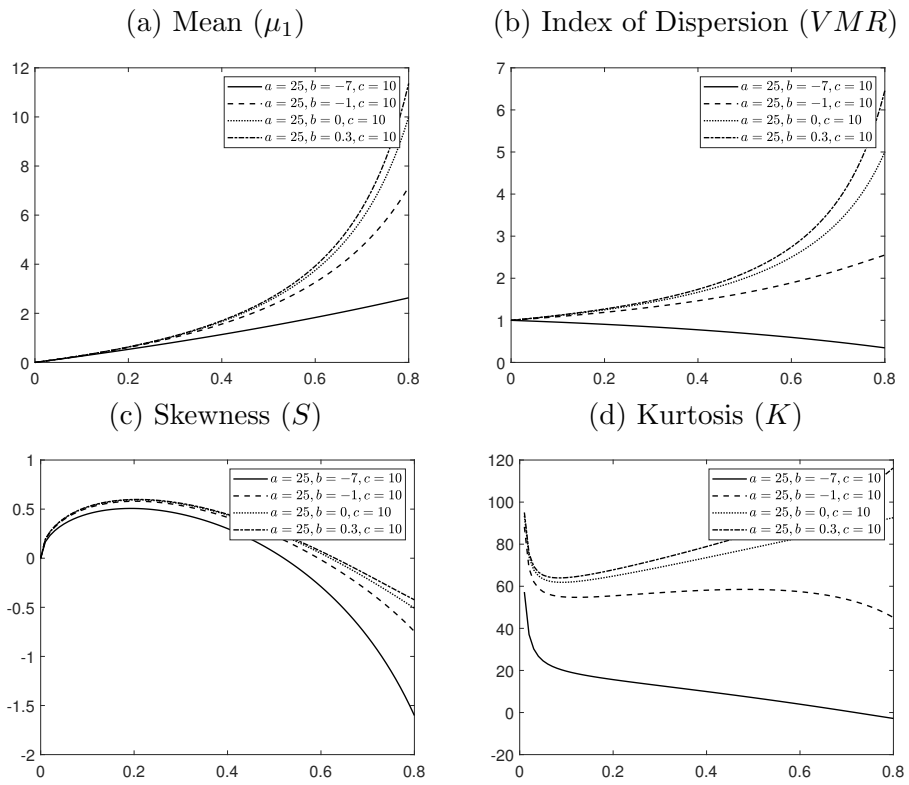


Figure 4.2: GLK moments when increasing the value of β (horizontal axis) for different values of b (lines).

$$K = (1 + 2b\theta - (b+c)c\theta^2)\left(\frac{(1-\beta-\beta^2)}{a\theta(1-\beta)} + \frac{5(b+c)}{\kappa}\right) + 3, \qquad (4.6)$$

respectively. For a given value of θ, there is negative skewness if $\beta < (1+\xi)/(2\kappa+\xi)$ with $\xi = 3\theta(b+c)a^{-1/2}$ and positive otherwise.

Figure 4.2 illustrates the effect of the parameter values on the mean, index of dispersion, skewness and kurtosis. Increasing the value of β (horizontal axis) the $\mathcal{GLK}(a, b, c, \beta)$ distribution allows for different types of dispersion (panel b), for both negative and positive skewness (panel c) and for various degrees of excess of kurtosis (panel d).

We conclude this section with an important property of the GLK distribution.

Proposition 2. *A random variable $X \sim \mathcal{GLK}(a, b, c, \beta)$ is infinite divisible, in particular $X \overset{\mathcal{L}}{=} \sum_{j=1}^{n} X_{jn}$ where $X_{jn} \overset{iid}{\sim} \mathcal{GLK}(a/n, b, c, \beta)$.*

See Appendix for a proof based on the properties of the pgf function.

4.2.2 A INAR(1) process

The Generalized Katz INAR(1) process (GLK-INAR(1)) is defined by means of the binomial thinning operator, $\circ$. The binomial thinning for a non-negative

discrete random variable X is defined as

$$\alpha \circ X = \sum_{i=1}^{X} B_i(\alpha) \tag{4.7}$$

where $B_i(\alpha)$ are iid Bernoulli r.v.s with success probability $P(B_i(\alpha) = 1) = \alpha$.

Definition 1 (GLK-INAR process)**.** For $\alpha \in (0, 1)$, the GLK-INAR(1) process is defined by

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \qquad t \in \mathbb{Z} \tag{4.8}$$

where ε_t are iid random variables with Generalized Lagrangian Katz distribution $\mathcal{GLK}(a, b, c, \beta)$, independent of X_s, $s \leq t - 1$.

Figure 4.3 provides some trajectories of $T = 100$ points each, simulated from a GLK-INAR(1) with innovation distributions given by the solid lines in the bottom plots of Fig. 4.1, that are $\mathcal{GLK}(3.86, 0, 0.60, 0.70)$ (overdispersion) and $\mathcal{GLK}(25.00, 0.00, 0.70, 0.42)$ (underdispersion). The trajectories correspond to two parameter settings we find the empirical application to climate change discussed in Section 4.4, that are: (i) high persistence setting ($\alpha = 0.7$, left); (ii) low persistence setting ($\alpha = 0.3$, right). In all plots, the empirical mean of the observations is reported (dashed line) as a reference to illustrate the different level of persistence in the trajectories.

Thanks to the general parametric family assumed, by setting $b = 0$, $c = \beta = \theta_1$ and $a = \theta_2$, our GLK-INAR(1) nests the INARKF(1) of Kim and Lee (2017) as special case. The GLK-INAR(1) naturally nests the Poisson INAR(1) of Al-Osh and Alzaid (1987), the Negative Binomial INAR(1) of Al-Osh and Aly (1992), and the Generalized Poisson INAR(1) of Alzaid and Al-Osh (1993).

As any INAR process, the GLK-INAR(1) has the following representation

$$X_{t+k} = \alpha^k \circ X_t + \sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t+k-j} \tag{4.9}$$

and its conditional pgf can be written as

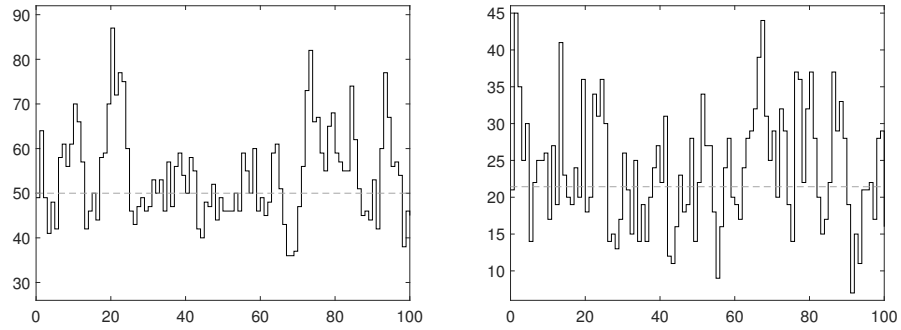
$$H_{X_{t+k}|X_t}(u) = (1 - \alpha^k + \alpha^k u)^{X_t} \prod_{j=0}^{k-1} H(1 - \alpha^j + \alpha^j u) \tag{4.10}$$

where $H(u)$ is defined in Eq. 4.3 or in Eq. 4.4. Details are given in Appendix 4.6.1. The previous representation allows us to derive the following properties, which are useful in forecasting.

Theorem 1. The conditional mean and variance of the process GLK-INAR(1) $\{X_t, t \in \mathbb{Z}\}$ are

$$\mathbb{E}(X_{t+k}|X_t) = \alpha^k X_t + \frac{1 - \alpha^{k-1}}{1 - \alpha} \frac{a\theta}{\kappa} \tag{4.11}$$

(a) Over-dispersed innovations



(a) Under-dispersed innovations

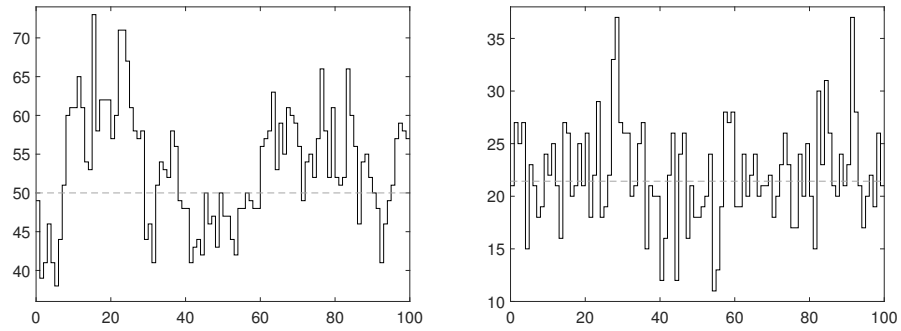


Figure 4.3: Trajectories of the GLK-INAR(1) in the high ($\alpha = 0.7$, left column) and low persistence ($\alpha = 0.3$, right column) regimes. The trajectories in the over- and under-dispersion settings are in the rows. In all plots, the dashed line is the empirical mean of the observations.

$$\mathbb{V}(X_{t+k}|X_t) = (\alpha^k - \alpha^{2k})X_t + \frac{1-\alpha^{2k}}{1-\alpha^2}\left(\frac{a(1-\beta)\theta}{\kappa^3} - \frac{a\theta}{\kappa}\right) + \frac{1-\alpha^k}{1-\alpha}\frac{a\theta}{\kappa} \quad (4.12)$$

where $\kappa = 1 - \beta - b\beta/c$ and $\theta = \beta/c$.

Remark 5. *Setting $b = 0$, $c = \beta = \theta_1$ and $a = \theta_2$ the results in Kim and Lee (2017) Th. 2.2 are obtained.*

Remark 6. *Since $\alpha < 1$, $\lim_{k\to\infty} \mathbb{E}(X_{t+k}|X_t) = a\theta/(\kappa(1-\alpha)$ and $\lim_{k\to\infty} \mathbb{V}(X_{t+k}|X_t) = a\theta((1-\beta) + \alpha\kappa^2)/((1-\alpha^2)\kappa^3)$ where $\kappa = 1 - \beta - b\beta/c$ and $\theta = \beta/c$.*

The process $\{X_t\}_{t\in\mathbb{Z}}$ is a Markov Chain on $\mathbb{N}$ and the transition probability $P_{i,j} = \mathbb{P}(X_t = j|X_{t-1} = i)$ can be expressed as

$$\begin{aligned} P_{i,j} &= \sum_{k=0}^{\min(i,j)} \mathbb{P}(\alpha \circ X_{t-1} = k|X_{t-1} = i)\mathbb{P}(\varepsilon = j-k) \\ &= \sum_{k=0}^{\min(i,j)} \binom{i}{k}\alpha^k(1-\alpha)^{i-k}p_{j-k} \end{aligned} \quad (4.13)$$

where p_x is the pmf given in Eq. 4.1.

Theorem 2. The process $\{X_t\}_{t\in\mathbb{Z}}$ is an irreducible, aperiodic and positive recurrent Markov chain. Hence there is a unique stationary distribution for the process $\{X_t\}_{t\in\mathbb{Z}}$.

See Appendix 4.6.1 for a proof. The result extends to the case of GLK innovations, the stationarity for INAR(1) with power series innovations given in Bourguignon and Vasconcellos (2015) and with Katz distribution given in Kim and Lee (2017).

From Th. 2 it follows that there is a strong stationary process with stationary distribution P_j, $j \in \mathbb{N}$, given by the marginal distribution

$$P_j = \sum_{i=0}^{\infty} P_{i,j} P_i. \tag{4.14}$$

Since at stationarity the process satisfies $X = \alpha \circ X + \varepsilon$, where $\varepsilon \sim \mathcal{GLK}(a, b, c, \beta)$, and the innovation terms are infinite divisible by Theorem 2, the stationary distribution satisfies the following definition of discrete semi-self-decomposability given in Bouzar (2008).

Definition 2. A nondegenerate distribution $\{p_x,\, x \in \mathbb{N}\}$ on $\mathbb{N}$ is said to be discrete semi-self-decomposable (DSSD) of order $\alpha \in (0, 1)$ if its pgf $\Psi(z)$ satisfies for all $|z| \leq 1$

$$\Psi(z) = \Psi(1 - \alpha - \alpha z)\Psi_\alpha(z)$$

where $\Psi_\alpha(z)$ is the pgf of an infinitely divisible distribution.

Being the stationary distribution a DSSD, Th. 2 in Bouzar (2008) yields that it is also infinitely divisible.

Theorem 3. The marginal distribution of the stationary process $\{X_t\}_{t\in\mathbb{Z}}$ is infinitely divisible.

Since the GLK distribution satisfies the convolution property (see Janardan, 1998, Th. 8), then the GLK-INAR(1) is stable by aggregation as stated in the following

Theorem 4. Let $\{X_{jt}\}_{t\in\mathbb{Z}}$ with $j = 1, 2, \ldots, J$ be a sequence of independent GLK-INAR(1) which satisfy:

$$X_{jt} = \alpha \circ X_{jt-1} + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim \mathcal{GLK}(a_j, b, c, \beta) \tag{4.15}$$

The process $Y_t = X_{1t} + \ldots + X_{Jt}$ is GLK-INAR(1) which satisfies:

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{GLK}(a_1 + \ldots + a_J, b, c, \beta) \tag{4.16}$$

Stationarity yields the following unconditional movements of the process.

Theorem 5. Let μ_ε, $\mu_\varepsilon^{(2)}$ and σ_ε^2 the mean, second order non-central moment and variance given in Prop. 1 for a $\mathcal{GLK}(a, b, c, \beta)$. For a GLK-INAR(1) process, the following unconditional moments can be derived:

(i) $\mu_X = \mathbb{E}(X_t) = \mu_\varepsilon/(1-\alpha)$

(ii) $\mu_X^{(2)} = \mathbb{E}(X_t^2) = (\alpha\mu_\varepsilon + 2\alpha\mu_\varepsilon^2/(1-\alpha) + \mu_\varepsilon^{(2)})/(1-\alpha^2)$

(iii) $\mathbb{E}(X_t X_{t-k}) = \alpha\mathbb{E}(X_{t-1}X_{t-k}) + \mu_\varepsilon\mu_X$

(iv) Higher order non-central moments can be derived using the formula:

$$\mu_X^{(m)} = \sum_{i=0}^{m}\sum_{k=0}^{i-1}\sum_{l=0}^{i-k}\binom{i}{k}(1-\alpha^i)^{-1}S(m,i)s(i-k,l)\alpha^k\mu_X^{(\underline{k})}\mu_\varepsilon^{(l)} \quad (4.17)$$

where $s(m,k)$ and $S(m,k)$ denote the Stirling's numbers of the I and II kind, respectively.

From the previous theorem one obtains the unconditional variance of the process $\sigma_X^2 = \mathbb{V}(X_t) = (\sigma_\varepsilon^2 + \alpha\mu_\varepsilon)/(1-\alpha^2)$ and the dispersion index of the process

$$VMR_X = \frac{\sigma_X^2}{\mu_X} = \frac{VMR_\varepsilon + \alpha}{1+\alpha} = \frac{1}{1+\alpha}\left(\alpha + \frac{1-\beta}{(1-\beta-b\beta/c)^2}\right) \quad (4.18)$$

where $VMR_\varepsilon = \sigma_\varepsilon^2/\mu_\varepsilon$ is the innovation index of dispersion. It follows that there is under- or over-dispersion in the marginal distribution, $VMR_X < 1$ and $VMR_X > 1$, if and only if there is under- or over-dispersion in the innovation, $VMR_\varepsilon < 1$ or $VMR_\varepsilon < 1$ respectively.

The autocorrelation function is

$$\gamma_k = \mathbb{C}ov(X_t, X_{t-k}) = \mathbb{E}(X_t X_{t-k}) - \mu_X^2 = \alpha^k\sigma_X^2 \quad (4.19)$$

as in the INAR(1) process (e.g., see Al-Osh and Alzaid, 1987).

4.3 Bayesian inference

4.3.1 Prior distribution

With the construction in Eq. 4.1 the constraint $\sum_{x\geq 0} p_x = 1$ is granted by the condition $H(1) = f(1) = 1$, nevertheless some constrains on the parameters are needed in order to have all the $p_x := \frac{1}{x!}\partial^{x-1}(g^x(z)f'(x))|_{x=0} > 0$. Three different cases are discussed below.

- For parameter values $a > 0, b \geq 0, c > 0$ the pmf are positive. Moreover, for $a/c, b/c \in \mathbb{N}$ the extended binomial coefficient $((a+bx)/c+1)_{x\uparrow}/x!$ coincides with the standard binomial coefficient $\binom{\frac{a+bx}{c}+x}{x}$ (Consul and Famoye, 2006, p. 8).

- For $-c < b < 0$, $a/c, b/c \in \mathbb{N}$ and $(c-a)/(c+b) \leq (a+c)/|b|$, the pmf are positive for $x < x^* = (a+c)/|b|$, while $p_x = 0$ for $x \geq x^*$.

- If $-c < b < 0$ but the additional constraints of the previous point are not satisfied, the terms appearing in the product $((a+bx)/c+1)_{x\uparrow}$ change sing and there is no guarantee that the result is positive. Indeed, for all the x such that $x > \max\{(a+1)/|b|, (c-a)/(c+b), 2\}$ one has $(a+bx)/c+1 < 0$ and $(a+bx)/c + x - 1 > 0$ and hence there is an integer $q = q_x$ such that $(a+bx)/c + m < 0$ for $1 \leq m \leq q$ and $(a+bx)/c + m > 0$ for $q+1 \leq m \leq k-1$. Hence

$$\frac{1}{\frac{a+bx}{c}+x}\Big(\frac{(a+bx)}{c}+1\Big)_{x\uparrow} = (-1)^q \prod_{m=1}^{q}\Big|\frac{(a+bx)}{c}+m\Big| \prod_{m=q+1}^{k-1}\Big|\frac{(a+bx)}{c}+m\Big|$$

  which is negative whenever q_x is odd. For example take $a = 10$, $b = -1$ and $c = 2$, for $x = 20$ one has $q_{20} = 5$, which shows that $p_{20} < 0$ which clearly is impossible.

In this paper we restrict our attention on the family of GLK and rule out negative pmf values by assuming $b \geq 0$. Our modelling and inference framework can be easily extend to account for these distributions.

Remark 7. *It should be noted that alternative definitions for $-c < b < 0$ can be considered. For example one can set to 0 the $p_x < 0$, i.e. when $x > \max\{(a+1)/|b|$. In this case re-scaling the p_x is necessary to get $\sum_{x=0}^{x^*} p_x = 1$. The resulting pmf is no more a generalized Lagrangian distribution (due to the truncation and rescaling) and the normalizing constant is not in closed form. See for example McCabe and Skeels (2020) for a discussion on the parameter values for the Katz distributions.*

In a Bayesian framework, the parameters constraints can be easily included in the inference process through a suitable choice of the prior distributions. We assume:

$$\alpha \sim \mathcal{B}e(\kappa_\alpha, \tau_\alpha), \quad a \sim \mathcal{G}a(\kappa_a, \tau_a), \quad b \sim \mathcal{G}a(\kappa_b, \tau_b), \tag{4.20}$$

$$c \sim \mathcal{G}a(\kappa_c, \tau_c), \quad \beta \sim \mathcal{B}e(\kappa_\beta, \tau_\beta) \tag{4.21}$$

where $\mathcal{B}e(\kappa, \tau)$ is the beta distribution with shape parameters κ and τ and $\mathcal{G}a(\kappa, \tau)$ the gamma distribution with shape and scale parameters κ and τ, respectively. In the empirical applications we assume a non-informative hyper-parameter setting for α and β, that is $\kappa_\alpha = \tau_\alpha = \kappa_\beta = \tau_\beta = 1$ and an informative prior for a, b and c with $\kappa_a = \tau_a = 1$, $\kappa_b = \kappa_c = 2$ and $\tau_b = \tau_c = 1/2$.

4.3.2 Posterior distribution

Let $x_1, \ldots, x_T$ be a sequence of observations for the GLK-INAR(1) process, then the joint posterior distribution is given by

$$\pi(\theta|x_1, \ldots, x_T) \propto \pi(\theta) \prod_{t=1}^{T} \prod_{i=0}^{\infty} \prod_{j=0}^{\infty} P_{ij}(\theta)^{\mathbb{I}(x_t-j)\mathbb{I}(x_{t-1}-i)} \tag{4.22}$$

where $\theta = (\alpha, a, b, c, \beta)$ is the parameter vector $\pi(\theta)$ the joint prior and

$$P_{ij}(\theta) = \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} \frac{\frac{a}{c}}{\frac{a+bx}{c} + j - k} \binom{\frac{a+b(j-k)}{c} + j - k}{j-k} \beta^{j-k} (1-\beta)^{\frac{a+b(j-k)}{c}}.$$

Remark 8 (Parameters space)**.** *Following the discussion above in this section, if the parameter constrain $c > 0$ is not imposed the coefficients of the Lagrangian expansion can be negative. In this case a truncated GLK can be used, similarly to what is proposed in McCabe and Skeels (2020) for the Katz distribution, and the inference procedure can be easily extended to include this type of distributions. The truncation can be imposed by using the following recursion for the transition probability:*

$$p_i(\theta) = p_0 \prod_{j=0}^{i-1} \max\left\{0, \frac{U(\theta) + V(\theta)j}{a+j}\right\} \tag{4.23}$$

where $U(\theta) = a\beta/c$, $V(\theta) = U(b+c)/(a+b)$ and

$$p_0 = \left(1 + \sum_{j=1}^{\infty} \prod_{k=0}^{j-1} \max\left\{0, \frac{U(\theta) + V(\theta)j}{a+j}\right\}\right)^{-1}. \tag{4.24}$$

The probability p_i becomes null for $i > j$ if $U(\theta) + V(\theta)j < 0$ at j.

Since the joint posterior is not tractable we follow a Markov Chain Monte Carlo (MCMC) framework to posterior approximation. See Robert and Casella (2013) for an introduction to MCMC methods. We overcome the difficulties in tuning the parameters of the MCMC procedure by applying the Adaptive MCMC sampler (AMCMC) proposed in Andrieu and Thoms (2008). Following a standard procedure, the following reparametrization is considered to impose the constrains to the parameters of the GLK-INAR(1). Let η be the 5-dimensional parameter vector obtained by the transformation $\varphi(\theta)$

$$\eta_1 = \log(\theta_1/(1-\theta_1)), \quad \eta_2 = \log(\theta_2) \tag{4.25}$$

$$\eta_3 = \log(\theta_3), \quad \eta_4 = \log(\theta_4), \quad \eta_5 = \log(\theta_5/(1-\theta_5)) \tag{4.26}$$

and let $\pi(\eta|x_1, \ldots, x_T) = \pi(\varphi^{-1}(\eta)|x_1, \ldots, x_T)J(\eta)$ be the posterior of η, with $J(\eta) = \theta_1\theta_2\theta_3\theta_4\theta_5(1-\theta_1)(1-\theta_5)$ the Jacobian of the transformation φ given above. Given the adaptation parameters μ^j and $\Sigma^{(j)}$, at the j-th iterations, the

AMCMC consists of the following three steps. First, a candidate η^* is generated from the following proposal distribution

$$\eta^* = \eta^{(j-1)} + \lambda^{(j)} w^{(j)}, \quad w^{(j)} \sim \mathcal{N}_q(\mathbf{0}, \Sigma^{(j)}). \tag{4.27}$$

Second, the candidate is accepted with probability $\rho^{(j)} = \rho(\eta^{(j-1)}, \eta^*)$, where

$$\rho(\eta^{(j-1)}, \eta^*) = \min\left(1, \frac{\pi(\varphi^{-1}(\eta^*)|x_1, \ldots, x_T) J(\eta^*)}{\pi(\varphi^{-1}(\eta^{(j-1)})|x_1, \ldots, x_T) J(\eta^{(j-1)})}\right) \tag{4.28}$$

and third, the adaptive parameters are updated as follows:

$$\begin{aligned}
\mu^{(j+1)} &= \mu^{(j)} + \gamma^{(j)}(\mu^{(j)} - \eta^{(j)}) && (4.29)\\
\Sigma^{(j+1)} &= \Sigma^{(j)} + \gamma^{(j)}((\mu^{(j)} - \eta^{(j)})(\mu^{(j)} - \eta^{(j)})' - \Sigma^{(j)}) && (4.30)\\
\log \lambda^{(j+1)} &= \log \lambda^{(j)} + \gamma^{(j)}(\rho^{(j)} - \rho^*) && (4.31)
\end{aligned}$$

where ρ^* is the target acceptance probability and $\gamma^{(j)} = j^{-a}$, $a > 0$ is the adaptive scale (Andrieu and Thoms, 2008, , Algorithm 4). Following the suggestions in Roberts et al. (1997) we set $\rho^* = 0.44$.

4.3.3 Simulation results

We illustrate the effectiveness of the Bayesian procedure in recovering the true value of the parameters and the efficiency of the MCMC procedure through some simulation experiments. We test the efficiency of the algorithm in two different settings, which can be commonly found in the data: low persistence and high persistence (see trajectories in Fig. 4.3). The true values of the parameters are: $\alpha = 0.3$, $a = 5.3239$,$b = 0.0592$, $c = 0.6$, $\beta = 0.5917$ in the low persistence setting, and $\alpha = 0.7$, $a = 5.3239$,$b = 0.0592$, $c = 0.6$, $\beta = 0.5917$ in the high persistence setting. For each setting we run the Gibbs sampler for 50,000 iterations on each dataset, discard the first 10,000 draws to remove dependence on initial conditions, and finally apply a thinning procedure with a factor of 10, to reduce the dependence between consecutive draws.

For illustrative purposes, we show in Figures 4.4 the MCMC posterior approximation for the parameter α (first row), the unconditional mean of the process (second row), and the marginal likelihood (last row), in one of our experiments for the high- and low-persistence settings. In each plot the true value (solid black line) and the Bayesian estimates approximated by using 4,000 MCMC samples after thinning and burn-in removal (dashed red line). For illustrative purposes, Fig. 4.10-4.11 and 4.12-4.13 in Appendix 4.6.2 exhibit 10,000 MCMC posterior draws and the MCMC approximation of the posterior distribution for all the parameters, in the high- and low-persistence settings.

In our experiments the acceptance rate is in the range of 40%-53% for both parameter settings. Table 4.1 shows, for all the parameters the autocorrelation function (ACF), effective sample size (ESS), inefficiency factor (INEFF)

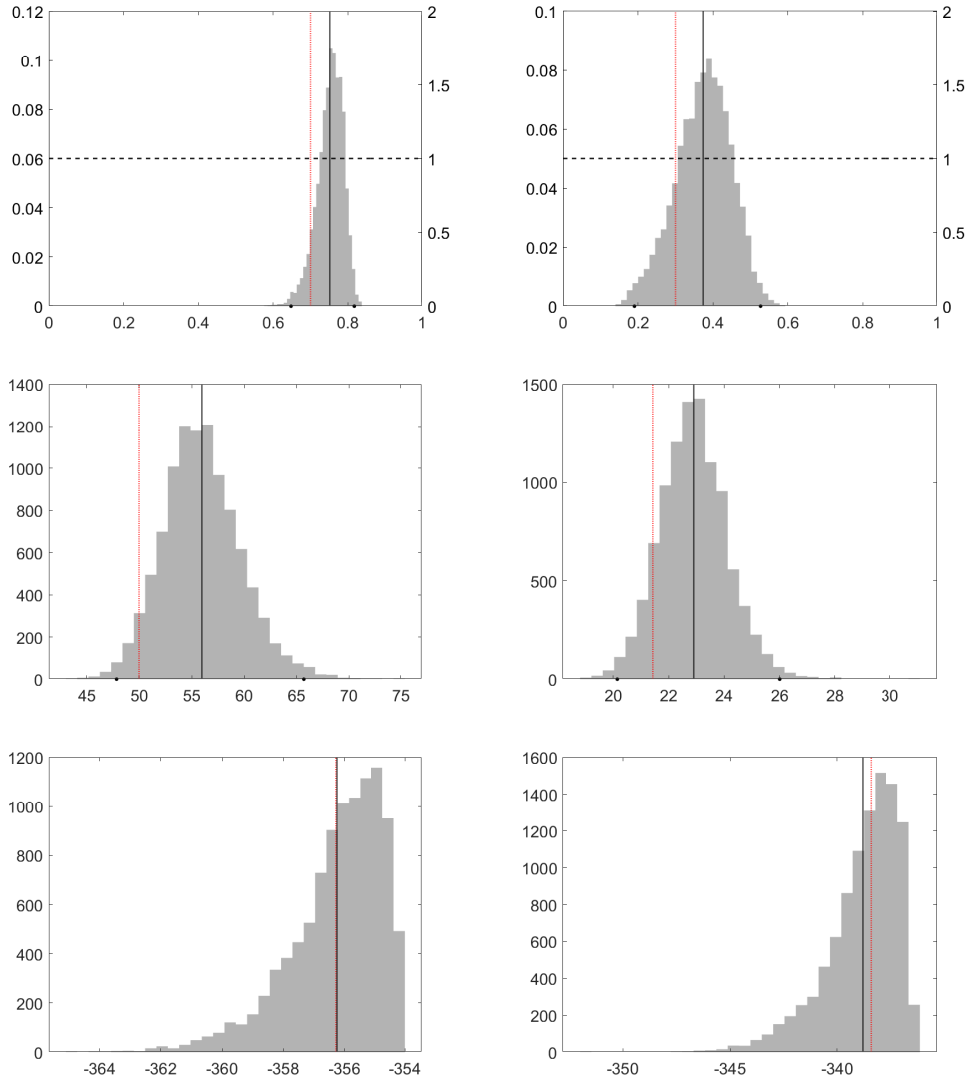


Figure 4.4: MCMC approximation of the posterior distribution (histogram) of the parameters α (top), the unconditional mean $\mu_\varepsilon/(1-\alpha)$ (middle) and the marginal likelihood (bottom) of the GLK-INAR(1) in the high-persistence (left) and low-persistence (right) setting. In all plots, the true parameter value (red dashed) and the estimated one (black solid).

	Low persistence ($\alpha = 0.3$, $a = 5.3239$,$b = 0.0592$, $c = 0.6$, $\beta = 0.5917$)					High persistence ($\alpha = 0.7$, $a = 5.3239$,$b = 0.0592$, $c = 0.6$, $\beta = 0.5917$)				
	α	a	b	c	β	α	a	b	c	β
$ACF(1)_{BT}$	0.93	0.92	0.93	0.92	0.94	0.94	0.92	0.93	0.92	0.94
$ACF(5)_{BT}$	0.71	0.68	0.72	0.66	0.74	0.72	0.68	0.73	0.65	0.74
$ACF(10)_{BT}$	0.52	0.47	0.54	0.45	0.56	0.53	0.48	0.55	0.44	0.55
$ACF(1)_{AT}$	0.54	0.47	0.55	0.44	0.58	0.53	0.48	0.56	0.45	0.57
$ACF(5)_{AT}$	0.08	0.06	0.13	0.08	0.16	0.06	0.05	0.11	0.02	0.08
$ACF(10)_{AT}$	-0.01	0.02	0.02	0.04	0.06	0.01	-0.003	-0.004	0.06	0.03
ESS_{BT}	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
ESS_{AT}	0.17	0.18	0.15	0.18	0.14	0.19	0.20	0.15	0.20	0.16
$INEFF_{BT}$	17.05	16.43	17.20	16.12	17.61	17.30	16.51	17.46	16.00	17.57
$INEFF_{AT}$	5.84	5.51	6.45	5.54	6.97	5.35	5.07	6.58	4.87	6.09
NSE_{BT}	0.002	0.05	0.002	0.006	0.002	0.001	0.06	0.004	0.01	0.004
NSE_{AT}	0.002	0.09	0.003	0.01	0.004	0.002	0.09	0.003	0.01	0.004
CD_{BT}	1.12	-2.08	-0.42	0.31	1.72	-0.48	-0.92	-0.13	-1.48	-0.83
	(0.26)	(0.04)	(0.68)	(0.76)	(0.09)	(0.63)	(0.36)	(0.89)	(0.14)	(0.41)
CD_{AT}	-1.047	0.57	-1.32	0.68	1.12	-1.05	0.57	-1.32	0.68	1.12
	(0.30)	(0.57)	(0.19)	(0.50)	(0.26)	(0.30)	(0.57)	(0.19)	(0.50)	(0.26)

Table 4.1: Autocorrelation function (ACF), effective sample size (ESS), inefficiency factor (INEFF), numerical standard errors (NSE) and Gewekes convergence diagnostic (CD) of the posterior MCMC samples for the two settings: low persistence and high persistence. We ran the proposed MCMC algorithm for 50,000 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 10. In parenthesis the p-values of the Geweke's convergence diagnostic.

and Geweke's convergence diagnostic (CD) before (BT subscript) and after thinning (AT subscript). The numerical standard errors are evaluated using the *nse* package (Geyer, 1992; Ardia and Bluteau, 2017; Ardia et al., 2018).

The thinning procedure is effective in reducing the autocorrelation levels and in increasing the ESS. 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 always accepted. The efficiency of the MCMC after the thinning procedure is generally improved. After thinning, on average the inefficiency measures (5.83), the p-values of the CD statistics (0.36) and the NSE (0.02) achieved the values recommended in the literature (e.g., see Roberts et al., 1997). Overall, this implies that the Gibbs sampler is computationally efficient, and the Monte Carlo estimator of the posterior quantities of interest has low variance.

4.4 Application to climate change

4.4.1 Data description

We used Google Trends data to measure the changes in public concern about climate change. Google Trend represents a source of big data (Choi and Varian, 2012; Scott and Varian, 2014) which have been used in many studies for example Anderberg et al. (2021) studied domestic violence during covid-19, Guolo and Varin (2014) and Yang et al. (2021) studied respectively flu and influenza trends, Schiavoni et al. (2021) and Yi et al. (2021) presented applications to unemployment while Yu et al. (2019) studied oil consumptions. In this study, we follow Lineman et al. (2015) and use Google search volumes as a proxy for public concern about "Climate Change" (CC) and "Global Warming" (GW). The search volume is the traffic for the specific combination of keywords relative to all queries submitted in Google Search in the world or in a given region of the world over a defined period. The indicator is in the range from 0 to 100, with 100 corresponding to the largest relative search volume during the period of interest. The search volume is sampled weekly from 4th December 2016 until 21st November 2021. We analysed the dynamics at the global and country level. Countries with an excess of zeros, above 95%, in the search volume series, have been excluded. The final dataset includes 65 countries of the about 200 countries provided by Google Trend. For illustration purposes, we report in the top plots of Fig. 4.5 the series of the world volume. The CC global volume exhibits overdispersion with $\widehat{VMR} = 102/27.33 = 3.73$, skewness and kurtosis $\widehat{S} = 2.09$ and $\widehat{K} = 13.47$, respectively. The GW global volume has over-dispersion $\widehat{VMR} = 170.42/48.56 = 3.51$, skewness $\widehat{S} = 0.27$ and kurtosis $\widehat{K} = 3.22$ (see also the histograms in the bottom plots). The country-specific indexes exhibit different levels of persistence and over-dispersion.

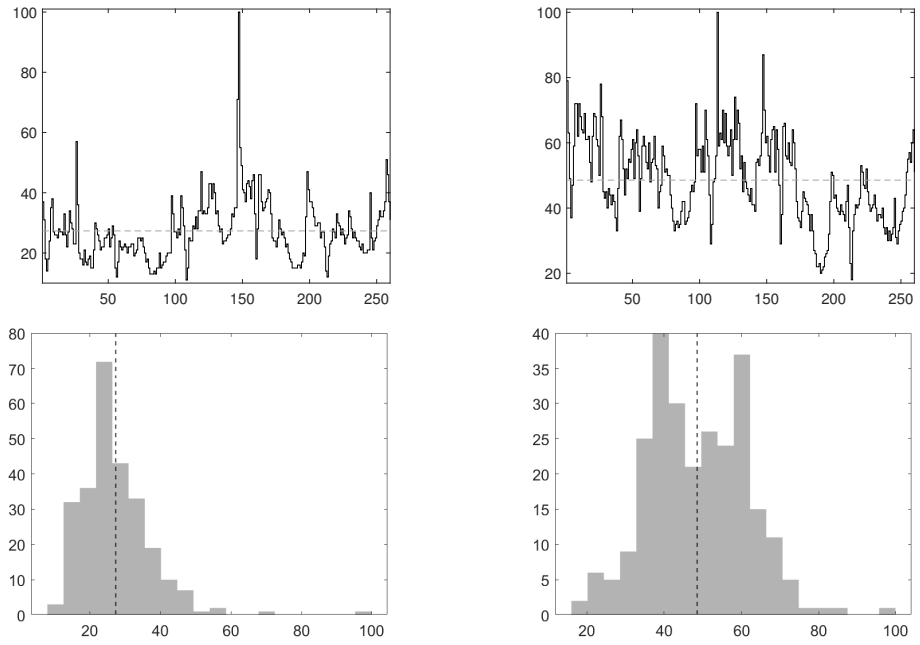


Figure 4.5: Time series (top) and histograms (bottom) of the global Google search of the words "Climate Change" (left) and "Global Warming" (right). Weekly frequency from 4th December 2016 to 21st November 2021. Empirical mean (dashed line).

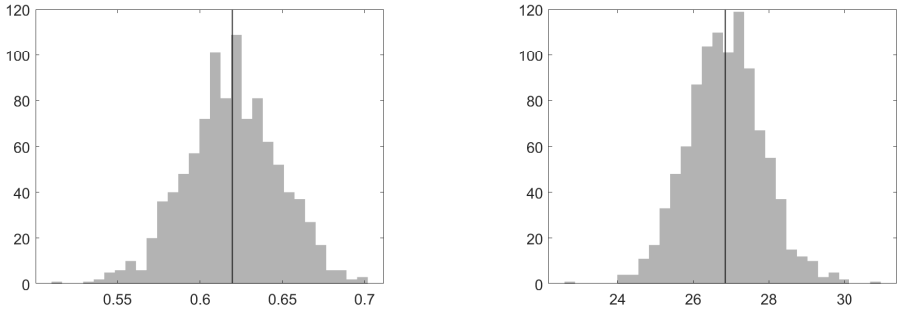
4.4.2 Estimation results

The posterior distribution of the autoregressive coefficient is given in Fig. 4.6. The coefficient estimate and posterior credible interval (in parenthesis) are $\widehat{\alpha} = 0.56$ $(0.50, 0.62)$ and $\widehat{\alpha} = 0.62$ $(0.56, 0.67)$ for the GW and the CC dataset, respectively. This result indicates that the public concern in climate risk is persistent over time worldwide at an aggregate level. The estimated parameter of the innovation process and their 0.95% credible intervals (in parenthesis) are $\widehat{a} = 3.53$ $(1.56, 6.08)$, $\widehat{b} = 0.04$ $(0.01, 0.11)$, $\widehat{c} = 0.21$ $(0.05, 0.47)$ and $\widehat{\beta} = 0.48$ $(0.20, 0.65)$ for the GW dataset and $\widehat{a} = 3.26$ $(1.44, 5.72)$, $\widehat{b} = 0.12$ $(0.021, 0.310)$, $\widehat{c} = 0.26$ $(0.032, 0.726)$ and $\widehat{\beta} = 0.35$ $(0.067, 0.623)$ for the CC one. The results indicate a deviation from the Negative Binomial model, thus we apply the DIC criterion $DIC = -4\mathbb{E}(\log p(X|\theta)|y) + 2\log p(X|\widehat{\theta})$ to compare GLK-INAR(1) and NB-INAR(1). The DIC is computed following (Spiegelhalter et al., 2002):

$$\text{DIC} = -4\frac{1}{N}\sum_{j=1}^{N}\log p(X|\theta^{(j)}) + 2\log p(X|\widehat{\theta}) \tag{4.32}$$

where $p(X|\theta)$ is the likelihood of the model, $\theta^{(j)}$ $j = 1, \ldots, N$ the MCMC draws after thinning and burn-in sample removal, and $\widehat{\theta}$ is the parameter estimate. The

(a) Google search dataset "Climate Change"



(b) Google search dataset "Global Warming"

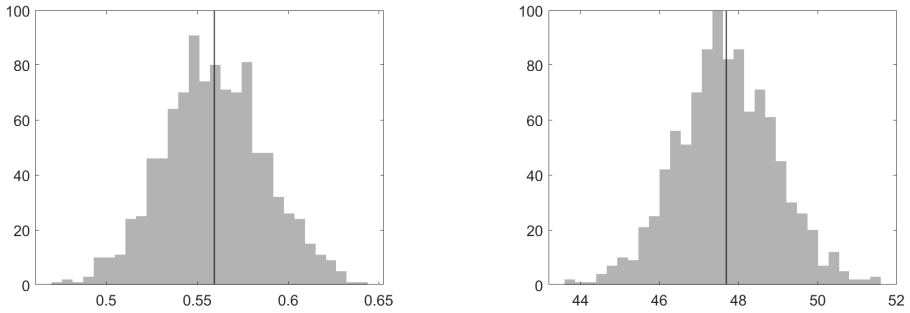


Figure 4.6: Posterior approximation of the persistence parameter α (left) and the unconditional moment $\mu_\varepsilon/(1-\alpha)$ (right) for the global search volume.

DICs for the GLK (NB) INARs fitted on the aggregate CC and GW series are $1.6743 \cdot 10^3$ ($1.6862 \cdot 10^3$) and $1.8735 \cdot 10^3$ ($1.8834 \cdot 10^3$), respectively.

We run the analysis at a disaggregate level. The results are given in Fig. 4.7-4.8 and Tab. 4.2-4.3. Figure 4.7 provides evidence of an inverse relationship between estimated persistence $\widehat{\alpha}$ and dispersion $\widehat{VMR}$ cross countries (reference lines in the left plot). There is evidence of this inverse relationship in both the CC (blue dots) and GW (red dots) datasets. The plot on the right indicates an inverse (direct) relationship between the estimated unconditional mean $\widehat{\mu}_\varepsilon/(1-\widehat{\alpha})$ and the dispersion index $\widehat{VMR}$ for the GW (CC). We indicate the parameter estimates for the world volume of searches (stars) in the same picture.

The terms "Climate Change" and "Global Warming" are used interchangeably, nevertheless, they describe different phenomena and can be used to determine the level of understanding of the public about these two parallel concepts Lineman et al. (2015). We investigate the relationships in the search volumes through the lens of our GLK-INAR(1) model. The left plot in Fig. 4.8 shows the unconditional mean of the search volumes for the two concepts in all countries (dots). In public attention, the two concepts are connected in the long run. We find a positive association for both countries with large (percentage of zeros $<$ 21%) and low search volumes (percentage of zeros $>$ 21%). There is an asymmetric effect in the overdispersion (right plot), and in all countries, the GW search

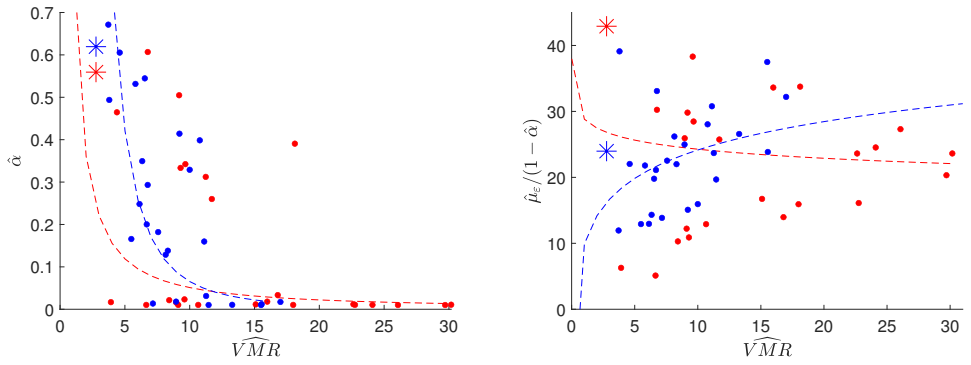


Figure 4.7: Persistence-dispersion ($\widehat{\alpha}$ and $\widehat{VMR}$, left) and unconditional mean and dispersion ($\widehat{\mu}_\varepsilon/(1-\widehat{\alpha})$ and $\widehat{VMR}$, right) scatter plots for all countries in the "Climate Change" (●) and "Global Warming" (●) datasets. Only countries with less than 21% of zeros are reported. Stars indicate the parameters of the world volume of searches. "*" indicates the parameter estimates for the aggregated search volume.

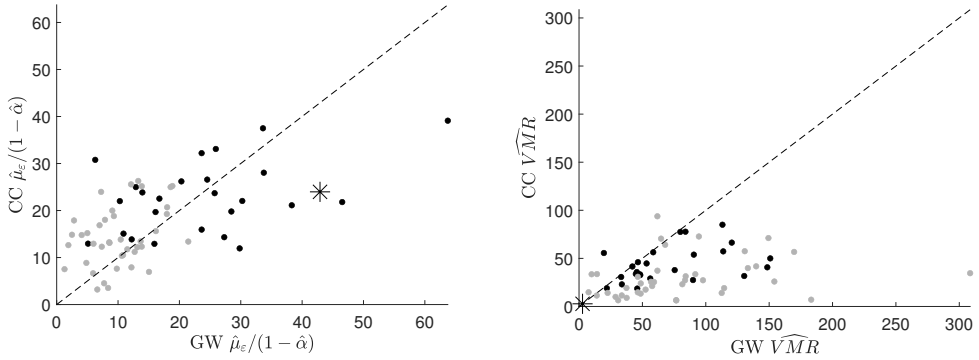


Figure 4.8: Unconditional mean (left) and dispersion index (right) of the GW (horizontal) and CC (vertical) for countries with more than 21% of zeros (●) less then 21% (•, values rescaled by five for visualization purposes) in the number of searches. In each plot the 45° reference line.

	Climate Change dataset				Climate Warming dataset			
Country	$\widehat{\alpha}$	CI	GLK	LK	$\widehat{\alpha}$	CI	GLK	LK
Australia	0.547	(0.501,0.589)	-882.43*	-885.87	0.343	(0.287,0.397)	-1125.99	-1120.90*
Bangladesh	0.018	(0.002,0.053)	-1118.68	-1111.99*	0.001	(0.001,0.005)	-1166.74	-1162.50*
Brazil	0.011	(0.001,0.029)	-1187.96	-1165.17*	0.001	(0.001,0.002)	-1027.29	-1024.77*
Canada	0.672	(0.615,0.714)	-711.55*	-716.70	0.512	(0.468,0.554)	-1003.35	-1000.60*
Emirates	0.010	(0.001,0.029)	-1200.72	-1182.74*	0.005	(0.001,0.020)	-1158.09	-1149.59*
France	0.184	(0.127,0.248)	-1069.66	-1068.36*	0.006	(0.001,0.019)	-1140.59	-1130.91*
Germany	0.298	(0.209,0.373)	-1023.87	-1022.87*	0.021	(0.001,0.060)	-1099.04	-1094.57*
India	0.499	(0.420,0.562)	-922.74*	-923.57	0.482	(0.417,0.541)	-1019.07	-1018.57*
Indonesia	0.021	(0.003,0.064)	-1123.06	-1108.15*	0.253	(0.193,0.311)	-1195.34	-1172.30*
Ireland	0.333	(0.279,0.388)	-953.35	-952.64*	0.001	(0.001,0.002)	-1093.97	-1094.39*
Italy	0.169	(0.090,0.241)	-983.83	-982.26*	0.001	(0.001,0.004)	-1036.21	-1034.28*
Malaysia	0.002	(0.001,0.008)	-1171.22	-1167.40*	0.006	(0.001,0.031)	-1135.02	-1129.73*
Mexico	0.009	(0.001,0.031)	-1160.49	-1148.85*	0.001	(0.001,0.001)	-1098.96	-1097.73*
Netherlands	0.148	(0.075,0.218)	-1064.29	-1061.30*	0.001	(0.001,0.005)	-1066.26	-1059.54*
NewZealand	0.353	(0.283,0.412)	-894.23*	-895.29	0.013	(0.001,0.044)	-1151.01	-1136.31*
Nigeria	0.129	(0.068,0.191)	-1043.77	-1041.80*	0.017	(0.001,0.062)	-974.18	-966.83*
Pakistan	0.212	(0.148,0.272)	-1131.60*	-1124.74	0.023	(0.001,0.064)	-1135.66	-1128.55*
Philippine	0.409	(0.354,0.460)	-1069.39	-1066.66*	0.383	(0.327,0.432)	-1150.86	-1138.26*
Singapore	0.159	(0.095,0.222)	-1099.02	-1094.02*	0.006	(0.001,0.023)	-1132.94	-1122.04*
SouthAfrica	0.413	(0.353,0.467)	-923.33*	-925.21	0.334	(0.274,0.391)	-865.23*	-868.48
Spain	0.253	(0.193,0.320)	-883.65*	-886.68	0.001	(0.001,0.002)	-1049.48	-1041.62*
Thailand	0.008	(0.001,0.035)	-1082.74	-1078.49*	0.004	(0.001,0.012)	-1218.44	-1193.67*
UK	0.535	(0.486,0.587)	-932.80*	-937.25	0.319	(0.246,0.388)	-1084.84	-1081.81*
US	0.601	(0.549,0.649)	-867.57*	-871.12	0.606	(0.558,0.649)	-941.39*	-942.22
Vietnam	0.003	(0.001,0.012)	-1189.79	-1178.82*	0.001	(0.001,0.001)	-987.59	-986.21*

Table 4.2: Estimated GLK-INAR(1) autoregressive coefficient ($\widehat{\alpha}$) and its 95% credible interval (CI), and marginal likelihood of the GLK-INAR(1) and the INARKF(1) models, for the "Climate Change" and "Global Warming" search volumes in different countries. Countries with less than 21% of zeros in the two series. "*" indicate the model with the largest marginal likelihood.

volume has a larger VMR than the CC volume. This can be explained by the GW larger variability induced by the changes in the use of the GW term in the official communications.

Comparing the coefficients across the rows of Tables 4.2-4.3, we find evidence of two types of series, one with high persistence and the other with low persistence. Moreover, for each country the level of persistence is similar across the two datasets (compare columns of Tables 4.2-4.3).

Tables 4.2-4.3 report the marginal likelihood of the GLK-INAR(1) and Lagrangian Katz INAR(1) in columns GLK and LK respectively. We find evidence of a better fitting of the GLK-INAR(1) for some countries and variables, e.g. CC searches in India and CC and GW searcher in South Africa. In order to get further insights into the results, we study the relationship between the dynamic and dispersion properties of the series and the actual level of climate risk of the countries. We consider the Global Climate Risk Index (CRI), which ranks countries and regions following the impacts of extreme weather events (such as storms, hurricanes, floods, heatwaves, etc.). The lower the index value, the larger the climate risk is. Following the values of the CRI for 2021, based on the events recorded from 2000 to 2019, our dataset includes some of the countries most

	Climate Change dataset				Climate Warming dataset			
Country	$\widehat{\alpha}$	CI	GLK	LK	$\widehat{\alpha}$	CI	GLK	LK
Argentina	0.001	(0.001,0.001)	-1022.22*	-1044.08	0.001	(0.001,0.001)	-803.30*	-850.10
Austria	0.001	(0.001,0.001)	-1085.17	-1082.50*	0.001	(0.001,0.001)	-717.08*	-754.18
Belgium	0.002	(0.001,0.012)	-1141.59	-1136.92*	0.001	(0.001,0.001)	-879.09*	-944.73
Colombia	0.001	(0.001,0.002)	-1040.35*	-1053.91	0.001	(0.001,0.001)	-848.15*	-882.50
Denmark	0.008	(0.001,0.023)	-1123.97	-1101.34*	0.002	(0.001,0.005)	-973.88	-950.20*
Egypt	0.001	(0.001,0.002)	-1103.36*	-1114.87	0.001	(0.001,0.002)	-855.84*	-867.85
Ethiopia	0.002	(0.001,0.011)	-1089.74	-1081.68*	0.001	(0.001,0.001)	-876.11*	-914.27
Finland	0.027	(0.001,0.084)	-987.30	-983.16*	0.001	(0.001,0.001)	-670.46*	-679.12
Ghana	0.003	(0.001,0.012)	-992.09	-980.8*3	0.001	(0.001,0.001)	-854.44*	-922.58
Jamaica	0.001	(0.001,0.005)	-995.13	-994.24*	0.001	(0.001,0.001)	-900.84*	-947.66
Greece	0.001	(0.001,0.001)	-1070.08*	-1095.69	0.001	(0.001,0.001)	-620.97*	-687.79
HongKong	0.005	(0.001,0.040)	-1116.20	-1104.70*	0.001	(0.001,0.001)	-1070.33*	-1076.02
Iran	0.001	(0.001,0.002)	-1046.05*	-1107.53	0.001	(0.001,0.002)	-945.79*	-992.93
Israel	0.001	(0.001,0.001)	-914.20*	-959.02	0.001	(0.001,0.001)	-726.50*	-797.23
Japan	0.008	(0.001,0.026)	-1209.16	-1193.53*	0.002	(0.001,0.004)	-1099.15*	-1109.15
Kenya	0.104	(0.049,0.170)	-1100.01	-1092.16*	0.001	(0.001,0.001)	-1073.55*	-1096.71
Lebanon	0.001	(0.001,0.001)	-761.11*	-776.44	0.001	(0.001,0.001)	-800.14*	-850.50
Morocco	0.001	(0.001,0.001)	-755.97*	-839.02	0.001	(0.001,0.002)	-471.85*	-534.09
Mauritius	0.001	(0.001,0.002)	-887.72*	-926.89	0.001	(0.001,0.001)	-601.12*	-655.86
Myanmar	0.001	(0.001,0.001)	-917.99*	-979.43	0.001	(0.001,0.002)	-602.14*	-665.81
Nepal	0.001	(0.001,0.001)	-1148.05	-1145.39*	0.001	(0.001,0.001)	-1027.36*	-1060.00
Norway	0.016	(0.003,0.051)	-1121.71	-1105.17*	0.001	(0.001,0.001)	-1002.83*	-1004.64
Peru	0.001	(0.001,0.001)	-915.67*	-950.87	0.001	(0.001,0.001)	-666.93*	-712.80
Polish	0.001	(0.001,0.001)	-1078.86*	-1090.00	0.001	(0.001,0.001)	-1000.76*	-1063.99
Portugal	0.002	(0.001,0.010)	-1035.62	-1030.57*	0.001	(0.001,0.001)	-800.35*	-851.13
Qatar	0.001	(0.001,0.001)	-879.41*	-917.09	0.001	(0.001,0.001)	-674.32*	-701.04
Romania	0.001	(0.001,0.001)	-880.49*	-901.20	0.001	(0.001,0.001)	-819.31*	-896.17
Russia	0.001	(0.001,0.003)	-1038.70*	-1050.47	0.001	(0.001,0.001)	-984.09*	-1023.75
StHelena	0.001	(0.001,0.001)	-873.70*	-914.40	0.001	(0.001,0.002)	-374.81*	-394.04
SouthKorea	0.004	(0.001,0.019)	-1149.02*	-1142.78	0.001	(0.001,0.003)	-1051.25*	-1057.01
SriLanka	0.001	(0.001,0.001)	-1086.31*	-1109.74	0.001	(0.001,0.003)	-842.37*	-863.17
Sweden	0.136	(0.067,0.205)	-1031.72	-1026.82*	0.001	(0.001,0.001)	-1078.30*	-1089.70
Swiss	0.028	(0.005,0.063)	-1055.02	-1046.17*	0.001	(0.001,0.001)	-835.78*	-878.17
Taiwan	0.001	(0.001,0.001)	-1074.92*	-1085.60	0.001	(0.001,0.001)	-794.34*	-815.63
TrinidadTobago	0.001	(0.001,0.001)	-920.81*	-955.62	0.001	(0.001,0.001)	-809.94*	-858.50
Turkey	0.004	(0.001,0.019)	-1095.99	-1091.08*	0.001	(0.001,0.001)	-1085.46*	-1091.38
Ukraine	0.001	(0.001,0.001)	-902.24*	-949.78	0.001	(0.001,0.002)	-709.80*	-745.82
Hungary	0.001	(0.001,0.001)	-935.84	-953.58*	0.001	(0.001,0.001)	-642.03*	-717.52
Zambia	0.001	(0.001,0.003)	-1075.30	-1053.65*	0.001	(0.001,0.001)	-746.07*	-789.78
Zimbabwe	0.001	(0.001,0.002)	-1130.16	-1128.73*	0.001	(0.001,0.002)	-658.24*	-690.68

Table 4.3: Estimated GLK-INAR(1) autoregressive coefficient ($\widehat{\alpha}$) and its 95% credible interval (CI), and marginal likelihood of the GLK-INAR(1) and the INARKF(1) models, for the "Climate Change" and "Global Warming" search volumes in different countries. Countries with more than 21% of zeros in the two series. "*" indicate the model with the largest marginal likelihood.

exposed to the climate risk, such as Japan, Philippines, Germany, South Africa, India, Sri-Lanka and Canada (Eckstein et al., 2021, see).

We study the relationship between public interest in climate related topics and the level of climate risk by fitting the GLK-INAR unconditional mean and the CV on the CRI of each country, that is

$$M_j = \xi + \zeta C_j + \eta_j, \quad \eta_j \overset{iid}{\sim} (0, \sigma_\eta^2) \quad j = 1, \ldots, J \tag{4.33}$$

where J is the number of countries, C_j is the country-specific CRI value, and

M_j is alternatively the unconditional mean $\mu(\theta_j)$ or the $\mathrm{CV}(\theta_j)$, which are both functions of the country-specific GLK-INAR parameter vector θ_j.

The Least Squares (LS) fitting of M is denoted by $\widehat{M}_{\mathrm{LS}}(\boldsymbol{\theta}, \mathrm{C}) = \hat{\xi}(\boldsymbol{\theta}) + \hat{\zeta}(\boldsymbol{\theta})C$ where $\hat{\xi}_{\mathrm{LS}}$ and $\hat{\zeta}_{\mathrm{LS}}$ are the LS estimator. They are functions of two arguments: the collection of unknown country-specific parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)$ and the collection of CRI observations $\mathbf{c} = (C_1, \ldots, C_J)$. For the easy of notation, we drop the argument $\mathbf{c}$.

It is possible to account for the parameter uncertainty when studying the relationship between public concern and CRI. Since $\widehat{M}_{\mathrm{LS}}(\boldsymbol{\theta}, C)$ is a deterministic transformation of $\boldsymbol{\theta}$ it has a distribution naturally induced by the posterior of $\boldsymbol{\theta}$. MCMC approximation of the posterior distribution can be used to evaluate the quantiles of $\widehat{M}_{\mathrm{LS}}(\boldsymbol{\theta}, C)$ for $C \in [0, 132]$. We report in all plots of Fig. 4.9 the posterior median of $\widehat{M}_{\mathrm{LS}}$ (solid line) together with the 95% credible interval for each value of C varying on a regular grid over $[0, 132]$ (shaded area).

The left plot in Fig. 4.9 shows unconditional mean against the CRI. There is evidence of a positive relationship between the public interest in climate-related topics and the actual level of climatic risk. Otherwise said, the lower the CRI level, the larger the Google search volumes are (see dashed lines). For example India has high risk (CRI equal to 7) and very high long-run level of public attention.

The right plot reports the coefficient of variation against the CRI for all countries in the "Climate Change" (blue) and "Global Warming" (red) datasets. The dashed lines represent linear regressions estimated on the data. There is evidence of a negative relationship between dispersion of the public concern and climatic risk; that is, in countries with more significant risk levels, the Google search volumes are less over-dispersed.

4.5 Conclusion

A novel integer-valued autoregressive process is proposed with Generalized Lagrangian Katz innovations. Theoretical properties of the model, such as stationarity, moments, and semi-self-decomposability are provided. A Bayesian approach to inference is proposed, and an efficient Gibbs sampling procedure has been proposed. The modeling framework is applied to a Google Trend dataset on the public concern about climate change for 65 countries. New evidence is provided about the long-run level of public attention, its persistence and dispersion in countries with low and high level of climate risk.

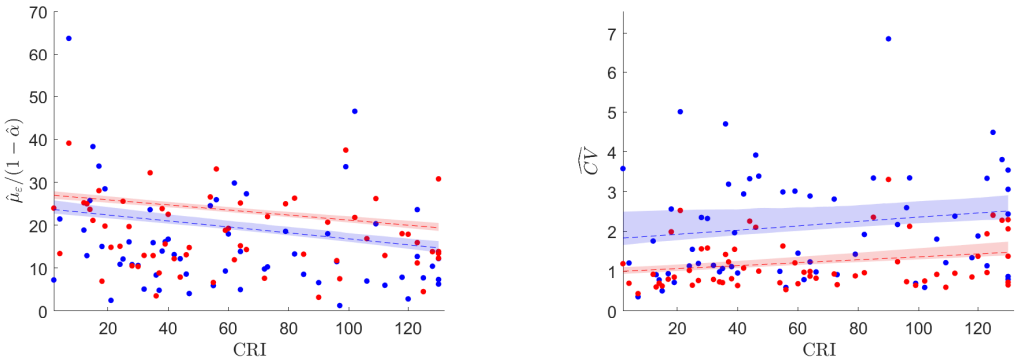


Figure 4.9: Climate Risk Index and unconditional mean scatter plot (CRI-$\mu_\varepsilon/(1-\alpha)$, left) and Climate Risk Index and dispersion scatter plot (CRI-CV, right) scatter plots for all countries in the "Climate Change" (●) and "Global Warming" (●) datasets. Dashed lines represent the linear regression estimated on the data and shaded areas provide the 95% credible regions.

Bibliography

Afrifa-Yamoah, E. and Mueller, U. (2022). Modeling digital camera monitoring count data with intermittent zeros for short-term prediction. *Heliyon*, 8(1):e08774.

Aknouche, A., Almohaimeed, B. S., and Dimitrakopoulos, S. (2021). Forecasting transaction counts with integer-valued garch models. *Studies in Nonlinear Dynamics & Econometrics*, page 20200095.

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.

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.

Anderberg, D., Rainer, H., and Siuda, F. (2021). Quantifying domestic violence in times of crisis: An internet search activity-based measure for the covid-19 pandemic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

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

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and computing*, 18(4):343–373.

Ardia, D. and Bluteau, K. (2017). nse: Computation of numerical standard errors in r. *Journal of Open Source Software*, 2(10):172.

Ardia, D., Bluteau, K., and Hoogerheide, L. F. (2018). Methods for computing numerical standard errors: Review and application to value-at-risk estimation. *Journal of Time Series Econometrics*, 10(2).

Battaglini, A., Barbeau, G., Bindi, M., and Badeck, F.-W. (2009). European winegrowers perceptions of climate change impact and options for adaptation. *Regional Environmental Change*, 9:6173.

Berry, L. R. and West, M. (2020). Bayesian forecasting of many count-valued time series. *Journal of Business & Economic Statistics*, 38(4):872–887.

Bourguignon, M., Vasconcellos, K. L., Reisen, V. A., and Ispny, M. (2016). A Poisson INAR(1) process with a seasonal structure. *Journal of Statistical Computation and Simulation*, 86(2):373–387.

Bourguignon, M. and Vasconcellos, K. L. P. (2015). First order non-negative integer valued autoregressive processes with power series innovations. *Brazilian Journal of Probability and Statistics*, 29(1):71 – 93.

Bouzar, N. (2008). Semi-self-decomposable distributions on $\mathbf{Z}_+$. *Annals of the Institute of Statistical Mathematics*, 60(4):901–917.

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. and Lee, S. (2017). Bayesian causality test for integer-valued time series models with applications to climate and crime data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(4):797–814.

Choi, H. and Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88:2–9.

Consul, P. C. and Famoye, F. (2006). *Lagrangian probability distributions*. Springer.

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.

Drovandi, C. C., Pettitt, A. N., and McCutchan, R. A. (2016). Exact and Approximate Bayesian Inference for Low Integer-Valued Time Series Models with Intractable Likelihoods. *Bayesian Analysis*, 11(2):325 – 352.

Eckstein, D., Künzel, V., and Schäfer, L. (2021). Global climate risk index 2021. Who suffers most from extreme weather events? Weather-related loss events in 2019 and 2000-2019. *Bonn: Germanwatch*.

Fahad, S. and Wang, J. (2018). Farmers risk perception, vulnerability, and adaptation to climate change in rural pakistan. *Land Use Policy*, 79:301–309.

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.

Fried, R., Agueusop, I., Bornkamp, B., Fokianos, K., Fruth, J., and Ickstadt, K. (2015). Retrospective Bayesian outlier detection in INGARCH series. *Statistics and Computing*, 25(2):365–374.

Frondel, M., Simora, M., and Sommer, S. (2017). Risk perception of climate change: Empirical evidence for Germany. *Ecological Economics*, 137:173–183.

Garay, A. M., Medina, F. L., Cabral, C. R., and Lin, T.-I. (2020a). Bayesian analysis of the p-order integer-valued ar process with zero-inflated poisson innovations. *Journal of Statistical Computation and Simulation*, 90(11):1943–1964.

Garay, A. M., Medina, F. L., Cabral, C. R. B., and Lin, T.-I. (2020b). Bayesian analysis of the p-order integer-valued ar process with zero-inflated poisson innovations. *Journal of Statistical Computation and Simulation*, 90(11):1943–1964.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science*, pages 473–483.

Guolo, A. and Varin, C. (2014). Beta regression for time series analysis of bounded data, with application to Canada Google Flu Trends. *The Annals of Applied Statistics*, 8(1):74 – 88.

Janardan, K. (1998). Generalized Polya Eggenberger family of distributions and its relation to Lagrangian Katz family. *Communications in Statistics-Theory and Methods*, 27:2423–2442.

Janardan, K. (1999). Estimation of parameters of the GPED. *Communications in Statistics-Theory and Methods*, 28:2167–2179.

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

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions, Third Edition*. John Wiley & Sons.

Katz, L. (1965). Unified treatment of a broad class of discrete probability distributions. *Classical and contagious discrete distributions*, 1:175–182.

Kim, H. and Lee, S. (2017). On first-order integer-valued autoregressive process with Katz family innovations. *Journal of Statistical Computation and Simulation*, 87(3):546–562.

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

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.

Lineman, M., Do, Y., Kim, J. Y., and Joo, G.-J. (2015). Talking about climate change and global warming. *PloS one*, 10(9):e0138996.

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

McCabe, B. P. and Skeels, C. L. (2020). Distributions you can count on but whats the point? *Econometrics*, 8(1):9.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 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.

Neal, P. and Subba Rao, T. (2007). MCMC for integer-valued ARMA processes. *Journal of Time Series Analysis*, 28(1):92–110.

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

Schiavoni, C., Palm, F., Smeekes, S., and van den Brakel, J. (2021). A dynamic factor model approach to incorporate big data in state space models for official statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(1):324–353.

Scott, S. L. and Varian, H. R. (2014). Predicting the present with Bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(1-2):4–23.

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

Shang, H. and Zhang, B. (2018). Outliers detection in INAR (1) time series. *Journal of Physics: Conference Series*, 1053:012094.

Singh, R. K., Zander, K. K., Kumar, S., Singh, A., Sheoran, P., Kumar, A., Hussain, S., Riba, T., Rallen, O., Lego, Y., Padung, E., and Garnett, S. T. (2017). Perceptions of climate variability and livelihood adaptations relating to gender and wealth among the adi community of the eastern indian himalayas. *Applied Geography*, 86:41–52.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

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

Ullah, H., Rashid, A., Liu, G., and Hussain, M. (2018). Perceptions of mountainous people on climate change, livelihood practices and climatic shocks: A case study of Swat District, Pakistan. *Urban Climate*, 26:244–257.

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

Weiß, C. H. and Kim, H.-Y. (2013). Parameter estimation for binomial AR(1) models with applications in finance and industry. *Statistical Papers*, 54(3):563–590.

Yang, S., Ning, S., and Kou, S. C. (2021). Use internet search data to accurately track state level influenza epidemics. *Scientific Reports*, 11(1):4023.

Yi, D., Ning, S., Chang, C.-J., and Kou, S. C. (2021). Forecasting unemployment using internet search data via prism. *Journal of the American Statistical Association*, 116(536):1662–1673.

Yu, L., Zhao, Y., Tang, L., and Yang, Z. (2019). Online big data-driven oil consumption forecasting with Google trends. *International Journal of Forecasting*, 35(1):213–223.

Ziegler, A. (2017). Political orientation, environmental values, and climate change beliefs and attitudes: An empirical cross country analysis. *Energy Economics*, 63:144–153.

4.6 Appendix

4.6.1 Proof of the results in Section 4.2

Proof of the resutls in Remark 1

Following (Consul and Famoye, 2006, p. 10-11) the general Lagrangian expansion of $f(u)$ is

$$\frac{f(z)}{1-zg'/g(z)} = \sum_{j=0}^{\infty} \frac{u^j}{j!} \left|\partial^j (g^j(z)f(z))\right|_{z=0} \tag{4.34}$$

where u satisfies $z = ug(z)$. The definition of Lagrangian distribution given in Janardan (1998) can is obtained from the one given above by replacing $f(z)$ with $f(z)(1 - zg'(z)/g(z)))$. By applying iteratively the derivative ∂ to the product of functions we obtain the coefficient in the j-th term of the expansion

$$\frac{1}{j!}|\partial^j \left(g(z)^j f(z)(1-zg'(z)/g(z))\right)|_{z=0} = \frac{1}{j!}|\partial^{j-1}(g(z)^j f'(z) \tag{4.35}$$

$$+(j-1)g'(z)g(z)^{j-1}f(z) - z\partial g^{j-1}(z)g'(z)f(z))|_{z=0} = \ldots \tag{4.36}$$

$$= \frac{1}{j!}|\partial^{j-1}(g(z)^j f'(z))|_{z=0} + |\partial^{j-\ell}\left((j-\ell)\partial^{\ell-1}(g'(z)g(z)^{j-1}f(z)\right) \tag{4.37}$$

$$-z\partial^{\ell}(g^{j-1}(z)g'(z)f(z))\big)\,|_{z=0} = \frac{1}{j!}|\partial^{j-1}(g(z)^j f'(z))|_{z=0} \tag{4.38}$$

where we set $\ell = j$ to get the result and the following equivalent Lagrangian expansion

$$\frac{f(z)}{1-zg'/g(z)} = \sum_{j=0}^{\infty} \frac{u^j}{j!} \left|\partial^j (g^j(z)f(z))\right|_{z=0} \tag{4.39}$$

$$\Leftrightarrow f(z) = \sum_{j=0}^{\infty} \frac{u^j}{j!} \left|\partial^{j-1} (g^j(z)f'(z))\right|_{z=0} \tag{4.40}$$

In particular, for $j \geq 1$,

$$p_j = \frac{1}{j!} \left|\partial^{j-1}(g^j(z)f'(z))\right|_{z=0}$$

provided that $p_j \geq 0$. Let $f(z) = \left(\frac{1-\beta}{1-\beta z}\right)^{a/c}$ and $g(z) = \left(\frac{1-\beta}{1-\beta z}\right)^{b/c}$ be the transformed and the transformer function, respectively. Then

$$f'(z) = \frac{a}{c}\left(\frac{1-\beta}{1-\beta z}\right)^{\frac{a}{c}+1} \frac{\beta}{1-\beta}, \tag{4.41}$$

$$g^k(z)f'(z) = \frac{a}{c}\left(\frac{1-\beta}{1-\beta z}\right)^{\frac{a}{c}+k\frac{b}{c}+1} \frac{\beta}{1-\beta}. \tag{4.42}$$

Hence

$$p_0 = f(0) = (1-\beta)^{\frac{a}{c}} \quad p_1 = g^1(0)f'(0) = \frac{a}{c}(1-\beta)^{\frac{a}{c}+\frac{b}{c}}\beta.$$

while, for $k \geq 2$, the k-th coefficient of the Lagrangian expansion in Eq. 4.39 is

$$\begin{aligned}
p_k &= \frac{1}{k!}|\partial^{k-1}(g(z))^k f'(z)|_{z=0} = \frac{1}{k!}\partial^{k-2}|(\frac{a}{c}\frac{\beta^2}{(1-\beta)^2}\xi_k\left(\frac{1-\beta}{1-\beta z}\right)^{\xi_k+1})|_{z=0} && (4.43)\\
&= \frac{1}{k!}|\partial^{k-3}(\frac{a}{c}\frac{\beta^3}{(1-\beta)^3}(\xi_k(\xi_k+1))\left(\frac{1-\beta}{1-\beta z}\right)^{\xi_k+2})|_{z=0} && (4.44)\\
&= ... && (4.45)\\
&= \frac{1}{k!}\beta^k\frac{a}{c}(1-\beta)^{\frac{a}{c}+k\frac{b}{c}}\prod_{m=0}^{k-2}(\xi_k+m) && (4.46)\\
&= \frac{1}{k!}\beta^k\frac{a}{c}(1-\beta)^{\frac{a}{c}+k\frac{b}{c}}\prod_{m=1}^{k-1}\left(\frac{a}{c}+k\frac{b}{c}+m\right) && (4.47)\\
&= \frac{1}{k!}\beta^k\frac{a}{c}(1-\beta)^{\frac{a}{c}+k\frac{b}{c}}\left(\frac{a}{c}+k\frac{b}{c}+1\right)_{k-1\uparrow} && (4.48)\\
&= \frac{1}{k!}\beta^k\frac{a}{c}\frac{1}{(\frac{a}{c}+k\frac{b}{c}+k)}(1-\beta)^{\frac{a}{c}+k\frac{b}{c}}\left(\frac{a}{c}+k\frac{b}{c}+1\right)_{k\uparrow} && (4.49)\\
& && (4.50)
\end{aligned}$$

where $\xi_k = \frac{a}{c}+k\frac{b}{c}+1$ and $(x)_{k\uparrow} = x(x+1)\dots(x+k-1)$ is the rising factorial.

- If $a>0, b>0, c>0$ one has $p_k > 0$ for every $k \geq 1$.
- If $-c < b < 0$, $a/c, b/c \in \mathbb{N}$ and $(c-a)/(c+b) \leq (a+c)/|b|$, then for $k < k^* = (a+c)/|b|$ one has

  $$\frac{a+bk}{c}+1 > 0$$

  and hence also

  $$\prod_{m=1}^{k-1}\left(\frac{a}{c}+k\frac{b}{c}+m\right) > 0$$

  proving that $p_k > 0$. For $k \geq k^* \geq (c-a)/(c+b)$ one has that $m_k = (|b|k-a)/c$ is an integer with $1 \leq m_k \leq k-1$ and hence the product $\prod_{m=1}^{k-1}\left(\frac{a}{c}+k\frac{b}{c}+m\right) = 0$ since for $m = m_k$ one has $\frac{a}{c}+k\frac{b}{c}+m = 0$. This shows that $p_k = 0$ for every $k \geq k^*$.

Proof of the results in Remark 3

Replacing c by β one obtains a Lagrangian Katz distribution. The LK is one of the few distributions which admits more pgfs. Let us consider the following

definition of pgf for a $\mathcal{LK}(a,b,\beta)$

$$G(u,a,b,\beta)=\left(\frac{1-\beta z}{1-\beta}\right)^{-\frac{a}{\beta}},\quad \text{with } z(a,b,\beta)=u\left(\frac{1-\beta z}{1-\beta}\right)^{-\frac{b}{\beta}}. \tag{4.51}$$

given in (Consul and Famoye, 2006, p. 241). Defining $n=(1-\beta z)/(\beta(z-1))$ and $1/\beta=n(z-1)+z$ the limiting pgf becomes

$$\lim_{\beta\to 0^+} G(u;a,b,\beta)=\lim_{n\to+\infty}\left(1+\frac{1}{n}\right)^{\frac{n(z-1)+z}{a}}=e^{a(z-1)}, \tag{4.52}$$

with

$$\lim_{\beta\to 0^+} z(a,b,\beta)=\lim_{n\to+\infty}\left(1+\frac{1}{n}\right)^{\frac{n(z-1)+z}{b}}=e^{b(z-1)} \tag{4.53}$$

which is the pgf of a GP given in (Consul and Famoye, 2006, pp. 166).

Proof of the results in Theorem 2

A random variable X is infinite divisible if for every $n\in\mathbb{N}$ there exists a sequence of random variables X_{nj}, $j=1,\ldots,n$ such that X has the same distribution of $X_{n1}+\ldots+X_{nn}$. From the pgf of a GLK given in Eq. 4.4

$$\mathbb{E}(X)=\mathbb{E}\left(u^X\right)=\left(\frac{1-\beta}{1-\beta z}\right)^{\frac{a}{c}}=\prod_{j=1}^{n}\left(\frac{1-\beta}{1-\beta z}\right)^{\frac{a}{nc}} \tag{4.54}$$

which is the pgf of the sum of n independent GLKs with distribution $\mathcal{GLK}(a/n,b,c,\beta)$ where $a/n>0$ according to the definition of GLK.

Proofs of Equations (4.10)**-**(4.9)

Before proving the general result let us prove the following two properties of the thinning operator

(i) $\alpha\circ(\alpha\circ X)=\alpha^2\circ X$

(ii) $\alpha\circ(X+Y)=\alpha\circ X+\alpha\circ Y$

where X and Y are two independent discrete-valued random variables. Property (i) follows from the law of iterated expectations and the properties of the probability generating function (pgf):

$$\mathbb{E}(\varphi^{\alpha\circ(\alpha\circ X)})=\mathbb{E}(\mathbb{E}(\varphi^{\alpha\circ(\alpha\circ X)}|\alpha\circ X))=\mathbb{E}(\prod_{j=1}^{\alpha\circ X}\mathbb{E}(\varphi^{B_j}|\alpha\circ X)) \tag{4.55}$$

$$=\mathbb{E}(((1-\alpha)+\varphi\alpha)^{(\alpha\circ X)})=((1-\alpha)+((1-\alpha)+\varphi\alpha)\alpha)^X \tag{4.56}$$

$$= (1 - \alpha^2 + \varphi\alpha^2)^X = \mathbb{E}(\varphi^{\alpha^2 \circ X}) \tag{4.57}$$

which is the pgf of the r.v. $\alpha^2 \circ X$. Property (ii) follows from:

$$\alpha \circ (X + Y) = \sum_{j=1}^{X+Y} B_j = \sum_{j=1}^{X} B_j + \sum_{j=X+1}^{X+Y} B_j \tag{4.58}$$

$$= \alpha \circ X + \sum_{l=1}^{Y} \tilde{B}_l = \alpha \circ X + \alpha \circ Y \tag{4.59}$$

where we defined $\tilde{B}_l = B_{l+X}$ for $l = 1, \ldots, Y$. Applying iteratively the two properties of the binomial thinning operator returns the following representation

$$X_{t+k} = \alpha \circ X_{t+k-1} + \varepsilon_{t+k} = \alpha^2 \circ X_{t+k-2} + \alpha \circ \varepsilon_{t+k-1} + \varepsilon_{t+k} \tag{4.60}$$

$$= \ldots = \alpha^k \circ X_t + \sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t+1-j} \tag{4.61}$$

where we defined $\alpha^0 \circ X = X$.

The conditional pgf follows from the above decomposition

$$H_{X_{t+k}|X_t}(u) = \mathbb{E}(u^{X_{t+k}}|X_t) = \mathbb{E}(u^{\alpha^k \circ X_t}|X_t) \prod_{j=0}^{k-1} \mathbb{E}(u^{\alpha^j \circ \varepsilon_{t+1-j}}) \tag{4.62}$$

$$= (1 - \alpha^k + \alpha^k u)^{X_t} \prod_{j=0}^{k-1} H(1 - \alpha^j + \alpha^j u) \tag{4.63}$$

where $H(u)$ satisfies Eq. 4.3 and 4.4.

Proof of the results in Theorem 1

Using the expressions for the moments given in Th. 1 and the independent assumption of the innovations on the thinning and on the past values of process one obtains for the case $k = 1$

$$\mathbb{E}(\alpha \circ X_t + \varepsilon_{t+1}|X_t) = \mathbb{E}(\sum_{i=1}^{X_t} B_{it} + \varepsilon_{t+1}|X_t) \tag{4.64}$$

$$= \alpha X_t + \mathbb{E}(\varepsilon_{t+1}) = \alpha X_t + \frac{a\theta}{\kappa} \tag{4.65}$$

and

$$\mathbb{V}(\alpha \circ X_t + \varepsilon_{t+1}|X_t) = \mathbb{V}(\sum_{i=1}^{X_t} B_{it}|X_t) + \mathbb{V}(\varepsilon_{t+1}) \tag{4.66}$$

$$= X_t(\alpha - \alpha^2) + \frac{a(1-\beta)\theta}{\kappa^3} \tag{4.67}$$

Following the recursive representation in 4.60 and by independence assumption, the conditional variance writes as

$$\mathbb{E}(\alpha \circ X_{t+k}|X_t) = \mathbb{E}(\alpha^k \circ X_t + \sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t+1-j}|X_t) \tag{4.68}$$

$$= \alpha^k X_t + \frac{1-\alpha^{k-1}}{1-\alpha}\frac{a\theta}{\kappa} \tag{4.69}$$

where we defined $\alpha^0 \circ X = X$. Similarly, the conditional variance writes as

$$\mathbb{V}(\alpha \circ X_{t+k}|X_t) = \mathbb{V}(\alpha^k \circ X_t + \sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t+1-j}|X_t) \tag{4.70}$$

$$= \mathbb{V}(\alpha^k \circ X_t|X_t) + \mathbb{V}(\sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t+1-j}) \tag{4.71}$$

$$= (\alpha^k - \alpha^{2k})X_t + \sum_{j=0}^{k-1}(\mathbb{V}(\mathbb{E}(\alpha^j \circ \varepsilon_{t+1-j}|\varepsilon_{t+1-j})) + \mathbb{E}(\mathbb{V}(\alpha^j \circ \varepsilon_{t+1-j}|\varepsilon_{t+1-j})) \tag{4.72}$$

$$= (\alpha^k - \alpha^{2k})X_t + \sum_{j=0}^{k-1}(\alpha^{2j}\frac{a(1-\beta)\theta}{\kappa^3} + (\alpha^j - \alpha^{2j}))\frac{a\theta}{\kappa}) \tag{4.73}$$

$$= (\alpha^k - \alpha^{2k})X_t + \frac{1-\alpha^{2k}}{1-\alpha^2}\left(\frac{a(1-\beta)\theta}{\kappa^3} - \frac{a\theta}{\kappa}\right) + \frac{1-\alpha^k}{1-\alpha}\frac{a\theta}{\kappa} \tag{4.74}$$

Proof of the result in Remark 5

Setting $b = 0$, $c = \beta = \theta_1$ and $a = \theta_2$, it follows $\kappa = 1 - \theta_2$ and $\theta = 1$, thus the conditional mean and variance become

$$\alpha^k X_t + \frac{1-\alpha^{k-1}}{1-\alpha}\frac{a\theta}{\kappa} = \alpha^k X_t + \frac{1-\alpha^{k-1}}{1-\alpha}\frac{\theta_1}{1-\theta_2} \tag{4.75}$$

and

$$(\alpha^k - \alpha^{2k})X_t + \frac{1-\alpha^{2k}}{1-\alpha^2}\left(\frac{a(1-\beta)\theta}{\kappa^3} - \frac{a\theta}{\kappa}\right) + \frac{1-\alpha^k}{1-\alpha}\frac{a\theta}{\kappa} \tag{4.76}$$

$$= (\alpha^k - \alpha^{2k})X_t + \frac{1-\alpha^{2k}}{1-\alpha^2}\left(\frac{\theta_1(1-\theta_2)}{(1-\theta_2)^3} - \frac{\theta_1}{(1-\theta_2)}\right) + \frac{1-\alpha^k}{1-\alpha}\frac{\theta_1}{(1-\theta_2)} \tag{4.77}$$

$$= (\alpha^k - \alpha^{2k})X_t + \frac{1-\alpha^{2k}}{1-\alpha^2}\frac{\theta_1}{(1-\theta_2)^2} + \frac{\theta_1}{(1-\theta_2)}\left(\frac{1-\alpha^k}{1-\alpha} - \frac{1-\alpha^{2k}}{1-\alpha^2}\right) \tag{4.78}$$

$$= (\alpha^k - \alpha^{2k})X_t + \frac{1-\alpha^{2k}}{1-\alpha^2}\frac{\theta_1}{(1-\theta_2)^2} + \frac{\theta_1}{(1-\theta_2)}\frac{(1-\alpha^k)(1+\alpha-1-\alpha^k)}{1-\alpha^2} \tag{4.79}$$

Proof of the result in Remark 6

Under the assumption $|\alpha| < 1$ the limiting conditional variance becomes:

$$\frac{1}{1-\alpha^2}\left(\frac{a(1-\beta)\theta}{\kappa^3} - \frac{a\theta}{\kappa} + \frac{a\theta(1+\alpha)}{\kappa}\right) \tag{4.80}$$

$$\frac{1}{1-\alpha^2}\left(\frac{a(1-\beta)\theta}{\kappa^3} - \frac{a\theta - a\theta(1+\alpha)}{\kappa}\right) \tag{4.81}$$

$$\frac{1}{1-\alpha^2}\left(\frac{a(1-\beta)\theta}{\kappa^3} + \frac{a\theta\alpha}{\kappa}\right) \tag{4.82}$$

$$\frac{1}{1-\alpha^2}\frac{a\theta}{\kappa^3}\left((1-\beta) + \alpha\kappa^2\right) \tag{4.83}$$

Proof of the results in Theorem 2

We follow the same argument as in Bourguignon and Vasconcellos (2015). Let $p_{l,k} = \mathbb{P}(X_t = k | X_{t-1} = l)$ be the transition probability given in Eq. 4.13. The process is irreducible since that every $k \in \mathbb{N}$ can be reached from every $l \in \mathbb{N}$. The transition are constant over t, thus the process admits a stationary distribution. We need to show that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{m=1}^{t} P_{0,0}^{m} > 0 \tag{4.84}$$

and bounded.

Next we show by induction that

$$P_{x,0}^{m} = p_0(1-\alpha^m)^x \prod_{j=1}^{m-1} H(1-\alpha^j) \tag{4.85}$$

where

$$H(1-\alpha^j) = \sum_{z=0}^{\infty} p_z(1-\alpha^j)^z. \tag{4.86}$$

For $m = 1$ the equation is satisfied since

$$P_{x,0}^{1} = p_0(1-\alpha)^x. \tag{4.87}$$

Assume it is satisfied for m then for $m+1$ we have

$$P_{x,0}^{m+1} = \sum_{z=0}^{\infty} P_{x,z}^{1} P_{z,0}^{m} \tag{4.88}$$

$$= \sum_{z=0}^{\infty} \sum_{k=0}^{\min(x,z)} \binom{x}{k} \alpha^k (1-\alpha)^{x-k} p_{z-k} P_{z,0}^{m} \tag{4.89}$$

$$= \sum_{k=0}^{x} \binom{x}{k} \alpha^k (1-\alpha)^{x-k} \sum_{z=k}^{\infty} p_{z-k} P^m_{z,0} \tag{4.90}$$

$$= \sum_{k=0}^{x} \binom{x}{k} \alpha^k (1-\alpha)^{x-k} \sum_{z=0}^{\infty} p_z P^m_{z+k,0} \tag{4.91}$$

$$= \sum_{k=0}^{x} \binom{x}{k} \alpha^k (1-\alpha)^{x-k} p_0 (1-\alpha^m)^k \sum_{z=0}^{\infty} p_z (1-\alpha^m)^z \tag{4.92}$$

$$= \sum_{k=0}^{x} \binom{x}{k} \alpha^k (1-\alpha)^{x-k} \sum_{z=0}^{\infty} p_z p_0 (1-\alpha^m)^{z+k} \prod_{j=1}^{m-1} H(1-\alpha^j) \tag{4.93}$$

$$= p_0 \prod_{j=1}^{m-1} H(1-\alpha^j) \sum_{k=0}^{x} \binom{x}{k} \alpha^k (1-\alpha)^{x-k} (1-\alpha^m)^k \sum_{z=0}^{\infty} p_z (1-\alpha^m)^z \tag{4.94}$$

$$= p_0 \prod_{j=1}^{m-1} H(1-\alpha^j) \sum_{k=0}^{x} \binom{x}{k} \alpha^k (1-\alpha)^{x-k} (1-\alpha^m)^k H(1-\alpha^m) \tag{4.95}$$

$$= p_0 \prod_{j=1}^{m} H(1-\alpha^j) \sum_{k=0}^{\min(x,z)} \binom{x}{k} \alpha^k ((1-\alpha^m)/(1-\alpha))^k (1-\alpha)^x \tag{4.96}$$

$$\tag{4.97}$$

where p_x is the pmf of the GLK and the summation in the last line is the pgf of the GLK. It follows that

$$P^{m+1}_{x,0} = p_0 \prod_{j=1}^{m} H(1-\alpha^j) \sum_{k=0}^{\min(x,z)} \binom{x}{k} \alpha^k ((1-\alpha^m)/(1-\alpha))^k (1-\alpha)^x \tag{4.98}$$

$$P^{m+1}_{x,0} = p_0 \prod_{j=1}^{m} H(1-\alpha^j) \sum_{k=0}^{\min(x,z)} \binom{x}{k} \alpha^k (1+\alpha+\ldots+\alpha^m)^k (1-\alpha)^x \tag{4.99}$$

$$= p_0 (1-\alpha)^x (1+\alpha+\ldots+\alpha^{m-1})^x \prod_{j=1}^{m} H(1-\alpha^j) \tag{4.100}$$

$$= p_0 (1-\alpha^{m+1})^x \prod_{j=1}^{m} H(1-\alpha^j) \tag{4.101}$$

Consider the logarithm

$$\log(P^{m+1}_{0,0}) = \log p_0 + \sum_{j=1}^{m} \log H(1-\alpha^j) = \log p_0 + \sum_{j=1}^{m} \log(1 + H(1-\alpha^j) - 1).$$

Now, recalling that $0 < \alpha < 1$, one has $0 \geq H(1-\alpha^j) - 1 \geq H(1-\alpha) - 1 =: -k_\alpha > -1$ for every $j \geq 1$. Since, by Taylor expansion, for every $|z| \leq k_\alpha < 1$,

$$|\log(1+z)| \leq |z| + C|z|^2$$

for a suitable constant $C = C(k_\alpha)$, then

$$|\log(P_{0,0}^{m+1})| \leq |\log p_0| + \sum_{j=1}^{m} \Big(|H(1-\alpha^j) - 1| + C|H(1-\alpha^j) - 1|^2 \Big).$$

Note that

$$1 - H(1-\alpha^j) = H(1) - H(1-\alpha^j) = (-\alpha^j)\dot{H}(\xi_\alpha)$$

for some $\xi_\alpha \in ((1-\alpha^j), 1)$. Now for any $x \in (0,1)$

$$\dot{H}(x) = \sum_{z \geq 1} z x^{z-1} p_z \leq \sum_{z \geq 1} z p_z =: M_1$$

since the GLK distribution has finte first moment. Hence one gets

$$|H(1-\alpha^j) - 1| \leq \alpha^j M_1$$

and

$$|\log(P_{0,0}^{m+1})| \leq |\log p_0| + \sum_{j \geq 1} \Big(M_1 \alpha^j + C M_1^2 \alpha^{2j} \Big) < +\infty$$

since $\alpha \in (0,1)$.

Remark 9. *We note that the proof only uses the fact that*

$$X_{t+1} = \alpha \circ X_t + \varepsilon_t$$

with $E[\varepsilon_t] < +\infty$.

Proof of the results in Theorem 5

(i) Under stationarity assumption one has $\mu_X = E(X_s)$ for all $s \in \mathbb{Z}$, thus $\mu_X = \alpha\mu_X + \mathbb{E}(\varepsilon_t)$ which implies $\mu_X = \mu_\varepsilon/(1-\alpha)$.
(ii) Let $\mu_X^{(2)} = E(X_s^2)$ for all $s \in \mathbb{Z}$, then $\mathbb{E}(X_t^2) = \mathbb{E}((\alpha \circ X_{t-1})^2) + \mathbb{E}(\varepsilon_t^2) + \mathbb{E}(2(\alpha \circ X_{t-1})\varepsilon_t) = \mathbb{V}((\alpha \circ X_{t-1})^2) + (\mathbb{E}(\alpha \circ X_{t-1}))^2 + \mathbb{E}(\varepsilon_t^2) + \mathbb{E}(2(\alpha \circ X_{t-1})\varepsilon_t)$. By the law of iterated expectation

$$\mu_X^{(2)} = \alpha^2(\mu_X^{(2)} - \mu_X^2) + \alpha(1-\alpha)\mu_X + \alpha^2\mu_X^2 + \mu_\varepsilon^{(2)} + \alpha\mu_X\mu_\varepsilon \quad (4.102)$$

$$\Leftrightarrow \mu_X^{(2)} = \frac{1}{1-\alpha^2}\left(\alpha\mu_\varepsilon + \mu_\varepsilon^{(2)} + \frac{2\alpha}{1-\alpha}\mu_\varepsilon^2\right) \quad (4.103)$$

From *(i)* and *(ii)* one can obtain the unconditional variance

$$\sigma_X^2 = \mu_X^{(2)} - \mu_X^2 \quad (4.104)$$

$$= \sigma_X^2 = \frac{1}{1-\alpha^2}\left(\alpha\mu_\varepsilon + \mu_\varepsilon^{(2)} + \frac{2\alpha}{1-\alpha}\mu_\varepsilon^2\right) - \frac{1}{(1-\alpha)^2}\mu_\varepsilon^2 \quad (4.105)$$

$$= \frac{1}{1-\alpha^2}\left(\alpha\mu_\varepsilon + \mu_\varepsilon^{(2)} + \frac{2\alpha}{1-\alpha}\mu_\varepsilon^2 - \frac{1+\alpha}{(1-\alpha)}\mu_\varepsilon^2\right) \quad (4.106)$$

$$= \frac{1}{1-\alpha^2}\left(\alpha\mu_\varepsilon + \sigma_\varepsilon^2\right) \tag{4.107}$$

From *(i)* and *(ii)* one can obtain the unconditional variance

$$\sigma_X^2 = \mu_X^{(2)} - \mu_X^2 \tag{4.108}$$
$$= \frac{1}{1-\alpha^2}\left(\alpha\mu_\varepsilon + \sigma_\varepsilon^2\right) - \frac{1}{(1-\alpha)^2}\mu_\varepsilon^2 \tag{4.109}$$
$$= \frac{1}{(1-\alpha^2)(1-\alpha)}\left(\alpha(1-\alpha)\mu_\varepsilon + (1-\alpha)\sigma_\varepsilon^2 - (1+\alpha)\mu_\varepsilon^2\right) \tag{4.110}$$
$$= \frac{1}{(1-\alpha^2)}\left(\alpha\mu_\varepsilon + \sigma_\varepsilon^2/(1-\alpha) - \mu_\varepsilon^2\right) \tag{4.111}$$
$$\tag{4.112}$$

(iii) $\mathbb{E}(X_tX_{t-k}) = \mathbb{E}((\alpha\circ X_{t-1}+\varepsilon_t)X_{t-k}) = \mathbb{E}(\mathbb{E}((\alpha\circ X_{t-1})X_{t-k}|X_{t-k},X_{t-1})) + \mathbb{E}(\varepsilon_t)\mathbb{E}(X_{t-k}) = \alpha\mathbb{E}(X_{t-1}X_{t-k}) + \mu_\varepsilon\mu_X$. From *(i)* and *(iii)* and stationarity $\mathbb{C}ov(X_t, X_{t-k}) = \gamma_k$ for all $t \in \mathbb{Z}$ one obtains the autocorrelation function

$$\mathbb{C}ov(X_t, X_{t-k}) = \mathbb{E}(X_tX_{t-k}) - \mathbb{E}(X_t)\mathbb{E}(X_{t-k}) \tag{4.113}$$
$$= \alpha\mathbb{E}(X_{t-1}X_{t-k}) + \mu_\varepsilon\mu_X - (\alpha\mathbb{E}(X_{t-1}) + \mu_\varepsilon)\mu_X \tag{4.114}$$
$$= \alpha\mathbb{C}ov(X_{t-1}X_{t-k}) = \alpha\gamma_k = \alpha^2\gamma_{k-2} = \cdots = \alpha^k\sigma_X^2 \tag{4.115}$$

(iii) Let us denote with $(x)_m = x(x-1)\dots(x-m+1)$ the falling factorial and with $\mu^{(\underline{k})} = \mathbb{E}((X)_k)$ the m-order falling factorial moment of a random variable X. The following two results will be used. The relationships between non-central moments and falling factorial moments are

$$\mathbb{E}(X_t^m) = \sum_{k=0}^{m} S(m,k)\mathbb{E}((X_t)_k) \tag{4.116}$$
$$\mathbb{E}((X_t)_m) = \sum_{k=0}^{m} s(m,k)\mathbb{E}(X_t^k) \tag{4.117}$$

where $s(m,k)$ and $S(m,k)$ are the Stirling numbers of the I and II kind, respectively (e.g., see Consul and Famoye, 2006, p. 18). Let X and Y be two random variables then

$$\mathbb{E}((X+Y)_m) = \sum_{k=0}^{m}\binom{m}{k}\mathbb{E}((X)_k)\mathbb{E}((X)_{m-k}) \tag{4.118}$$

which can be proved by induction. Let $\alpha\circ X$ a binomial thinning with X a discrete random variable, then

$$\mathbb{E}((\alpha\circ X)_k) = \mathbb{E}((\sum_{j=1}^{X} B_j)_k) = \mathbb{E}\left(\sum_{|\kappa|=k}\prod_{j=1}^{X} B_j^{\kappa_j}\right) = \mathbb{E}\left(\binom{X}{k}k!\alpha^k\right) = \alpha^k\mathbb{E}((X)_k) \tag{4.119}$$

where $|\kappa| = \kappa_1 + \ldots + \kappa_X$. Using the results given above and stationarity (i.e. $\mathbb{E}((X_t)_m) = \mu_X^{(\underline{m})}$ one obtains

$$\mathbb{E}((X_t)_m) = \mathbb{E}((\alpha \circ X_{t-1} + \varepsilon_t)_m) \tag{4.120}$$

$$= \sum_{k=0}^{m} \binom{m}{k} \mathbb{E}\left((\alpha \circ X_{t-1})_k\right) \mathbb{E}((\varepsilon_t)_{m-k}) \tag{4.121}$$

$$= \sum_{k=0}^{m} \binom{m}{k} \alpha^k \mathbb{E}((X_{t-1})_k) \mu_\varepsilon^{(\underline{m-k})} \tag{4.122}$$

which implies the m-order falling factorial moment of a INAR(1) is (see also).

$$\mu_X^{(\underline{m})} = \frac{1}{1-\alpha^m} \sum_{k=0}^{m-1} \binom{m}{k} \alpha^k \mu_X^{(\underline{k})} \mu_\varepsilon^{(\underline{m-k})} \tag{4.123}$$

$$= \frac{1}{1-\alpha^m} \sum_{k=0}^{m-1} \sum_{l=0}^{m-k} \binom{m}{k} s(m-k,l) \alpha^k \mu_X^{(\underline{k})} \mu_\varepsilon^{(l)} \tag{4.124}$$

and the m-order moment is

$$\mu_X^{(m)} = \sum_{i=0}^{m} S(m,i) \frac{1}{1-\alpha^i} \sum_{k=0}^{i-1} \sum_{l=0}^{i-k} \binom{i}{k} s(i-k,l) \alpha^k \mu_X^{(\underline{k})} \mu_\varepsilon^{(l)} \tag{4.125}$$

For example the second moment ($m = 2$) is

$$\mu_X^{(\underline{2})} = (1-\alpha^2)^{-1} \sum_{k=0}^{1} \binom{2}{k} \alpha^k \mu_X^{(\underline{k})} \mu_\varepsilon^{(\underline{2-k})} \tag{4.126}$$

$$= (1-\alpha^2)^{-1} (\mu_\varepsilon^{(2)} - \mu_\varepsilon + 2\alpha\mu_X\mu_\varepsilon) \tag{4.127}$$

where we use $\mu_\varepsilon^{(\underline{2})} = -\mu_\varepsilon + \mu_\varepsilon^{(2)}$ from Eq. 4.116. The third moment ($m = 3$) is

$$\mu_X^{(\underline{3})} = \frac{1}{1-\alpha^3} \left(\mu_\varepsilon^{(\underline{3})} + 3\alpha\mu_X\mu_\varepsilon^{(\underline{2})} + 3\alpha^2 \mu_X^{(\underline{2})} \mu_\varepsilon \right) \tag{4.128}$$

$$= \frac{1}{1-\alpha^3} \left(\mu_\varepsilon^{(\underline{3})} + 3\alpha\mu_X\mu_\varepsilon^{(\underline{2})} + 3\alpha^2 \mu_X^{(\underline{2})} \mu_\varepsilon \right) \tag{4.129}$$

$$= \frac{1}{1-\alpha^3} \left(\mu_\varepsilon^{(\underline{3})} + \frac{3\alpha^2\mu_\varepsilon}{1-\alpha^2} \left(\mu_\varepsilon^{(\underline{2})} + \frac{2\alpha\mu_\varepsilon^2}{1-\alpha^2} \right) + \frac{3\alpha\mu_\varepsilon\mu_\varepsilon^{(\underline{2})}}{1-\alpha} \right) \tag{4.130}$$

$$= \varphi(\alpha) \left(\left(2\mu_\varepsilon - 3\mu_\varepsilon^{(2)} + \mu_\varepsilon^{(3)} - 6\mu_\varepsilon\mu_\varepsilon^{(2)} + 6\mu_\varepsilon^2 + 6\mu_\varepsilon^3 \right) \alpha^3 \right. \tag{4.131}$$

$$+ \left(3\mu_\varepsilon^{(2)} - 3\mu_\varepsilon - \mu_\varepsilon^{(3)} + 3\mu_\varepsilon\mu_\varepsilon^{(2)} - 3\mu_\varepsilon^2 \right) \alpha^2 \tag{4.132}$$

$$\left. + \left(\mu_\varepsilon - \mu_\varepsilon^{(3)} + 3\mu_\varepsilon\mu_\varepsilon^{(2)} + 3\mu_\varepsilon^2 \right) \alpha + \mu_\varepsilon^{(3)} \right) \tag{4.133}$$

$$\tag{4.134}$$

where we set $\varphi(\alpha) = ((1-\alpha^3)(1-\alpha^2)(1-\alpha))^{-1}$ and used $\mu_\varepsilon^{(\underline{2})} = -\mu_\varepsilon + \mu_\varepsilon^{(2)}$ and $\mu_\varepsilon^{(\underline{3})} = 2\mu_\varepsilon - 3\mu_\varepsilon^{(2)} + \mu_\varepsilon^{(3)}$ from Eq. 4.116. Following similar argument as above the fourth moment ($m = 4$) is

$$\mu_X^{(\underline{4})} = \frac{1}{1-\alpha^4}\left(\mu_\varepsilon^{(\underline{4})} + 7\alpha\mu_X\mu_\varepsilon^{(\underline{3})} + 6\alpha^2\mu_X^{(\underline{2})}\mu_\varepsilon^{(\underline{2})} + \alpha^3\mu_X^{(\underline{3})}\mu_\varepsilon\right) \tag{4.135}$$

$$= \frac{1}{1-\alpha^4}\left(\mu_\varepsilon^{(\underline{4})} + \frac{\alpha^3\mu_\varepsilon}{1-\alpha^2}\left(3\alpha^2\mu_\varepsilon(\mu_\varepsilon^{(\underline{2})} + \frac{2\alpha\mu_\varepsilon^2}{1-\alpha}) - \mu_\varepsilon^{(\underline{3})} - \frac{3\alpha\mu_\varepsilon\mu_\varepsilon^{(\underline{2})}}{1-\alpha}\right)\right. \tag{4.136}$$

$$\left.\frac{1}{1-\alpha^2}\left(6\alpha^2\mu_\varepsilon^{(\underline{2})}(\mu_\varepsilon^{(\underline{2})} + \frac{2\alpha\mu_\varepsilon^2}{1-\alpha}\right) + \frac{1}{1-\alpha}\left(7\alpha\mu_\varepsilon\mu_\varepsilon^{(\underline{3})}\right)\right) \tag{4.137}$$

$$= \varphi(\alpha)\left(\left(\mu_\varepsilon\mu_\varepsilon^{(3)} - 3\mu_\varepsilon\mu_\varepsilon^{(2)} - 6\mu_\varepsilon^2\mu_\varepsilon^{(2)} + 2\mu_\varepsilon^2 + 6\mu_\varepsilon^3 + 6\mu_\varepsilon^4\right)\alpha^6\right. \tag{4.138}$$

$$+ \left(3\mu_\varepsilon\mu_\varepsilon^{(2)} - \mu_\varepsilon\mu_\varepsilon^{(3)} + 3\mu_\varepsilon^2\mu_\varepsilon^{(2)} - 2\mu_\varepsilon^2 - 3\mu_\varepsilon^3\right)(\alpha^5 + \alpha^4) \tag{4.139}$$

$$+ \left(11\mu_\varepsilon^{(2)} - 6\mu_\varepsilon - 6\mu_\varepsilon^{(3)} + \mu_\varepsilon^{(4)} + 30\mu_\varepsilon\mu_\varepsilon^{(2)} - 6\mu_\varepsilon\mu_\varepsilon^{(3)} + 12\mu_\varepsilon^2\mu_\varepsilon^{(2)}\right. \tag{4.140}$$

$$\left.-18\mu_\varepsilon^2 - 12\mu_\varepsilon^3 - 6(\mu_\varepsilon^{(2)})^2\right)\alpha^3 \tag{4.141}$$

$$+ \left(6\mu_\varepsilon^2 - 12\mu_\varepsilon\mu_\varepsilon^{(2)} + 6\mu_\varepsilon + 6(\mu_\varepsilon^{(2)})^2 - 11\mu_\varepsilon^{(2)} + 6\mu_\varepsilon^{(3)} - \mu_\varepsilon^{(4)}\right)\alpha^2 \tag{4.142}$$

$$+ \left(6\mu_\varepsilon - 11\mu_\varepsilon^{(2)} + 6\mu_\varepsilon^{(3)} - \mu_\varepsilon^{(4)} - 21\mu_\varepsilon\mu_\varepsilon^{(2)} + 7\mu_\varepsilon\mu_\varepsilon^{(3)} + 14\mu_\varepsilon^2\right)\alpha \tag{4.143}$$

$$\left.-6\mu_\varepsilon + 11\mu_\varepsilon^{(2)} - 6\mu_\varepsilon^{(3)} + \mu_\varepsilon^{(4)}\right) \tag{4.144}$$

where we used $\mu_\varepsilon^{(\underline{3})} = 2\mu_\varepsilon - 3\mu_\varepsilon^{(2)} + \mu_\varepsilon^{(3)}$ and $\mu_\varepsilon^{(\underline{4})} = -6\mu_\varepsilon + 11\mu_\varepsilon^{(2)} - 6\mu_\varepsilon^{(3)} + \mu_\varepsilon^{(4)}$ from Eq. 4.116.

4.6.2 Further simulation results

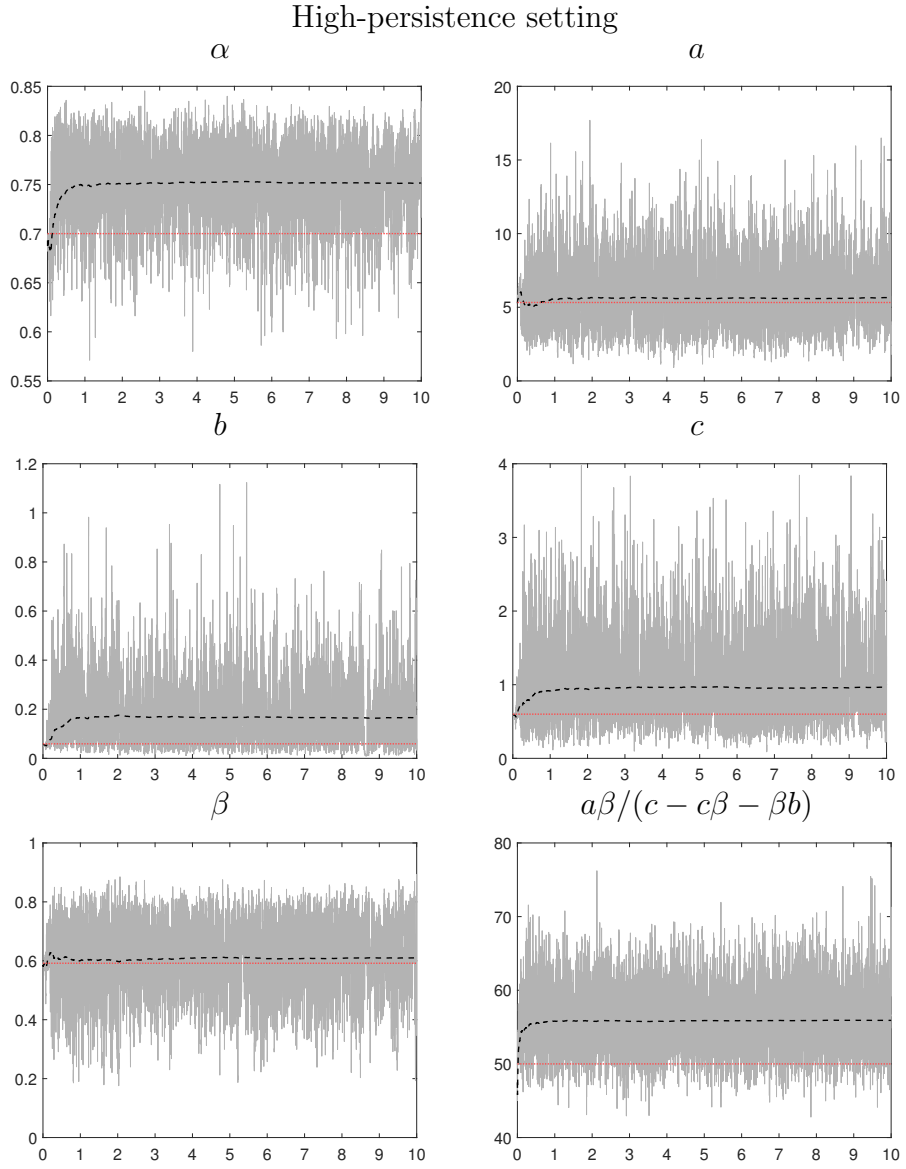


Figure 4.10: MCMC output for the parameters of the GLK-INAR(1). In all plots, the MCMC draws (gray solid), the progressive MCMC average (dashed black) over the iterations (horizontal axis in thousands), and the true value of the parameter (horizontal red dashed).

4.6.3 Further real data results

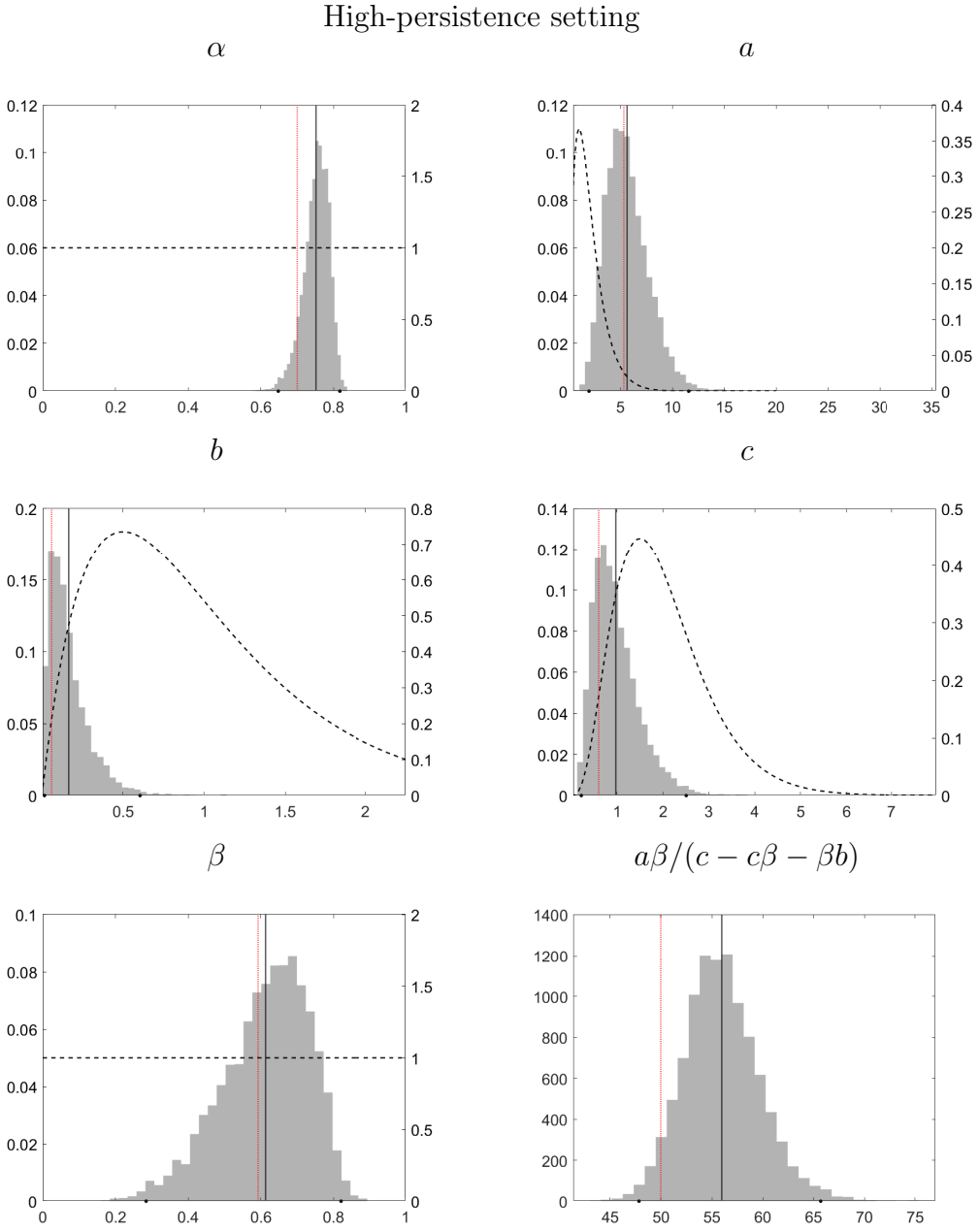


Figure 4.11: MCMC approximation of the posterior distribution (histogram) of the parameters. In all plots, the estimated value (vertical black solid), the true value (vertical red dotted) and the prior density (dashed).

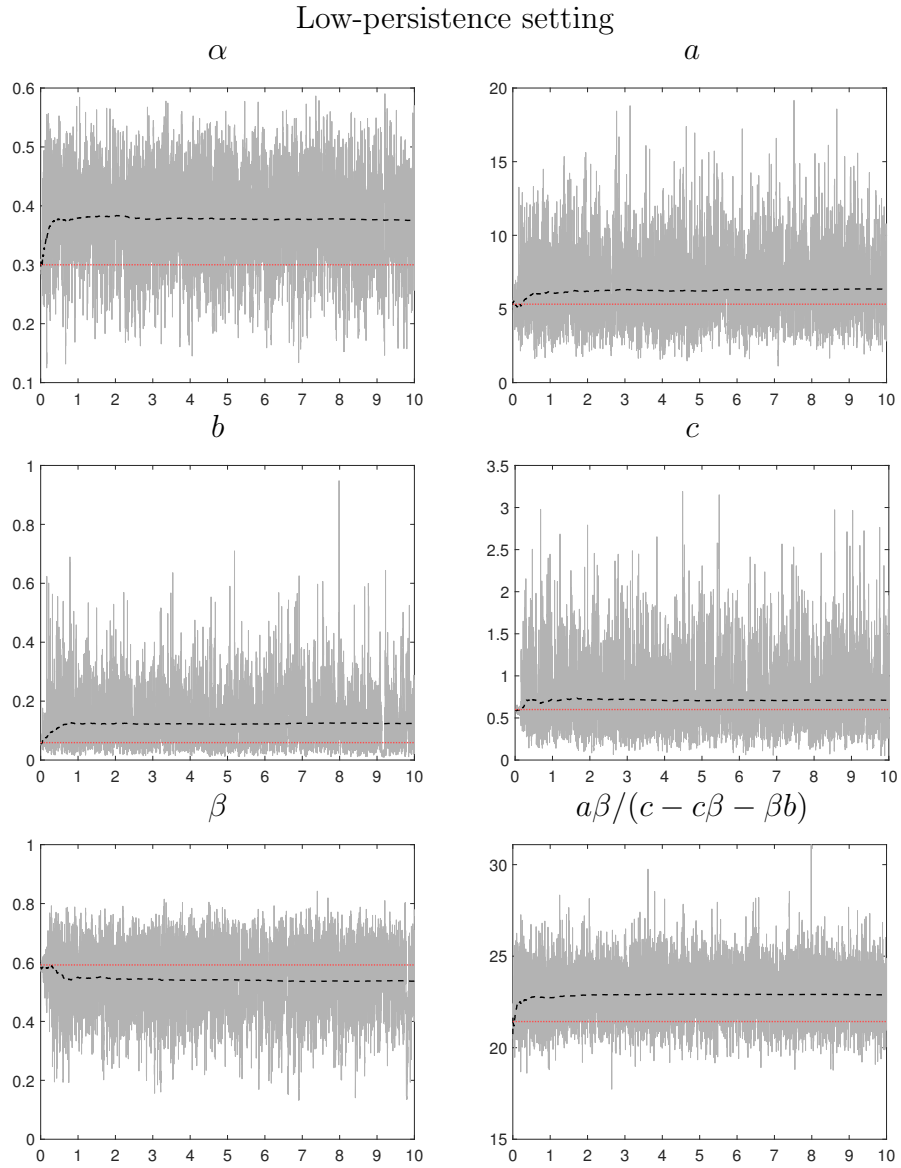


Figure 4.12: MCMC output for the parameters of the GLK-INAR(1). In all plots, the MCMC draws (gray solid), the progressive MCMC average (dashed black) over the iterations (horizontal axis in thousands), and the true value of the parameter (horizontal red dashed).

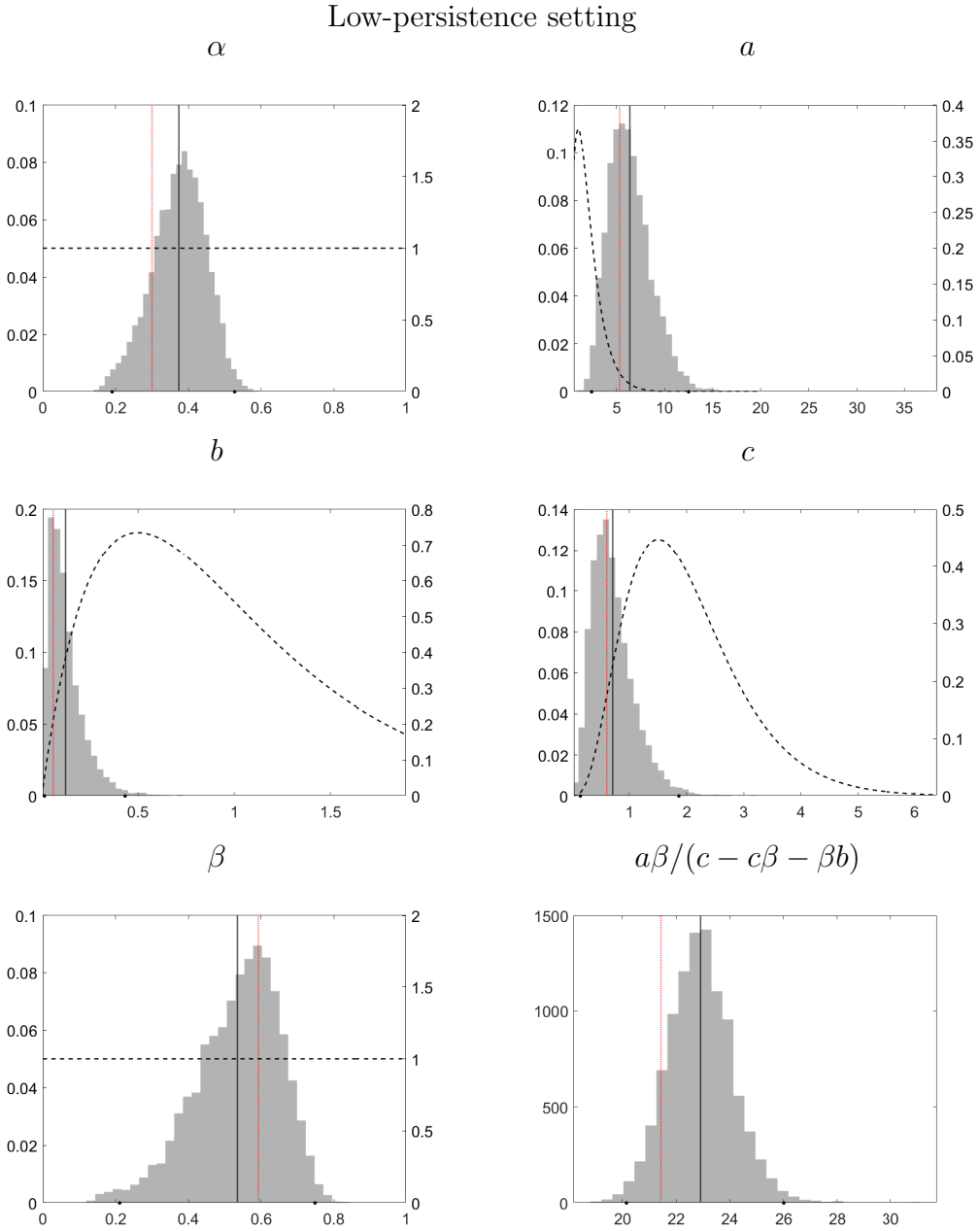


Figure 4.13: MCMC approximation of the posterior distribution (histogram) of the parameters. In all plots, the estimated value (vertical black solid), the true value (vertical red dotted) and the prior density (dashed).

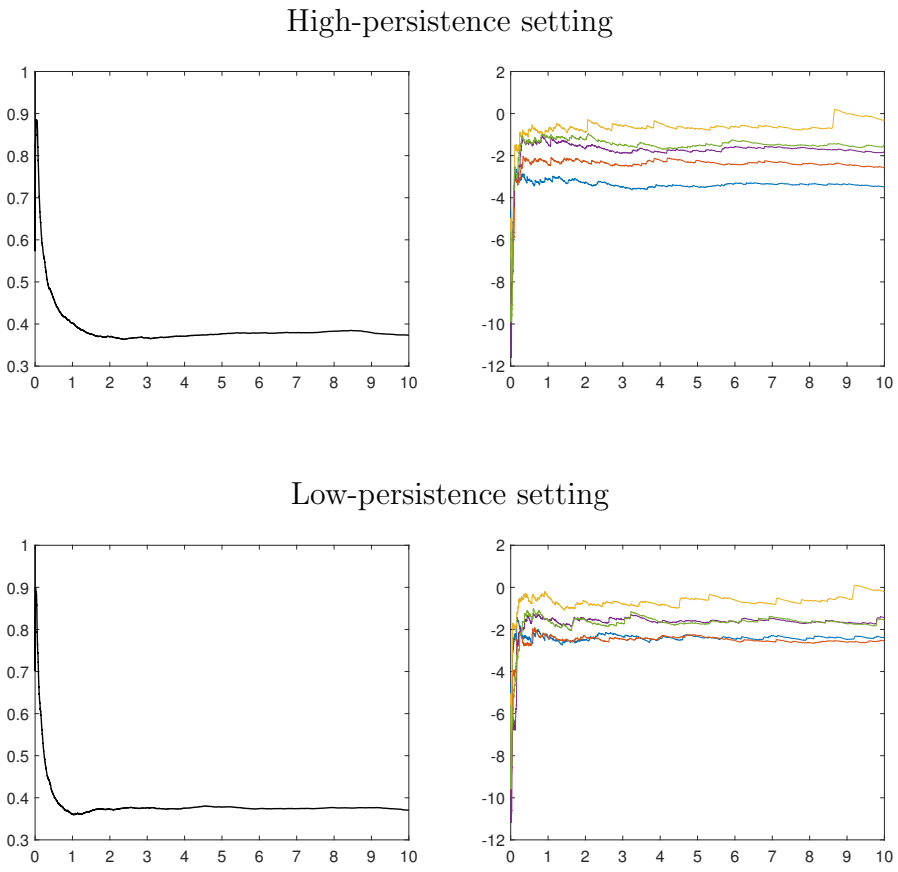


Figure 4.14: MCMC acceptance rate (left) and adaptive log-scales (right) over the iterations (horizontal axis in thousands).

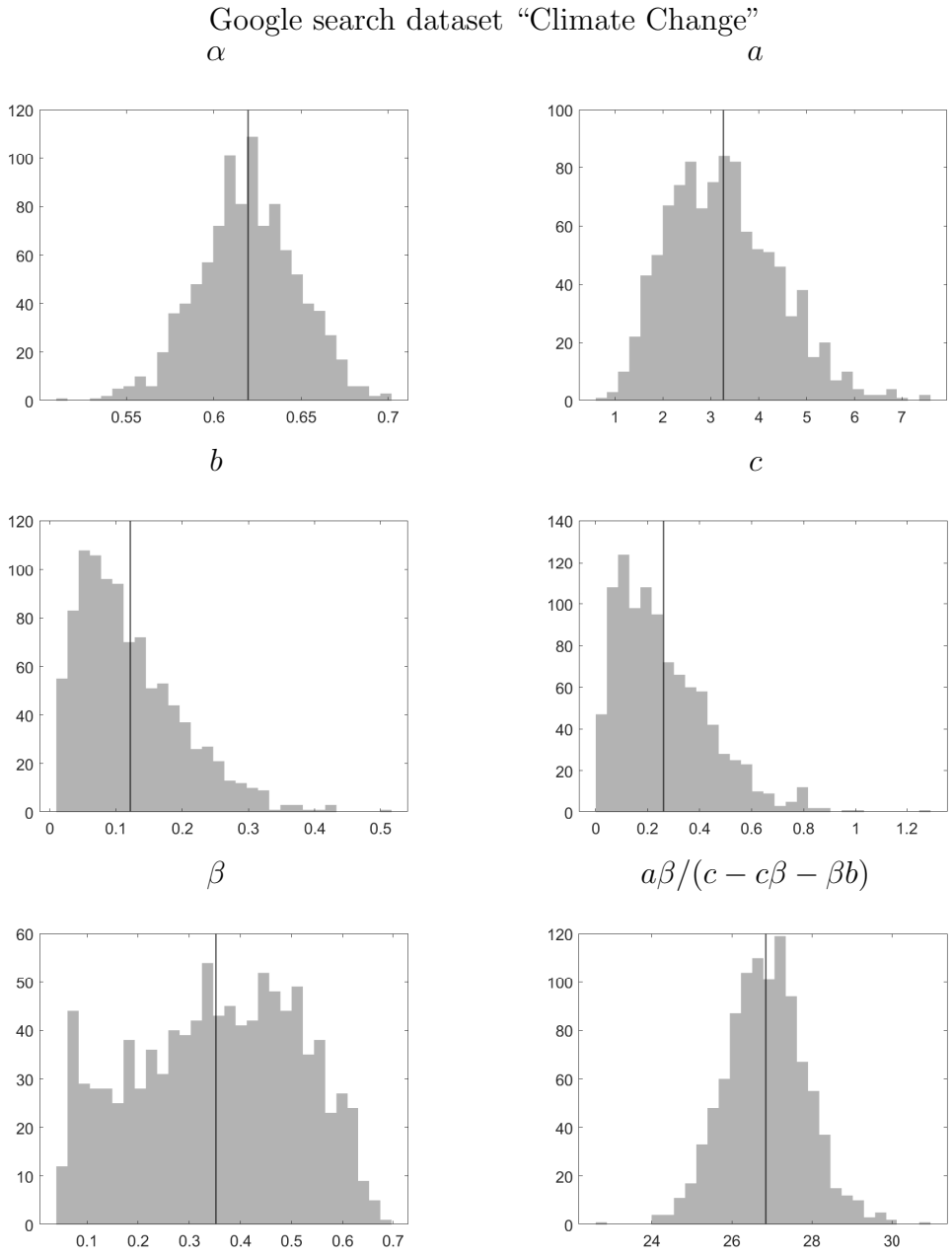


Figure 4.15: MCMC approximation of the posterior distribution (histogram) of the parameters. In all plots, the estimated value (vertical black solid).

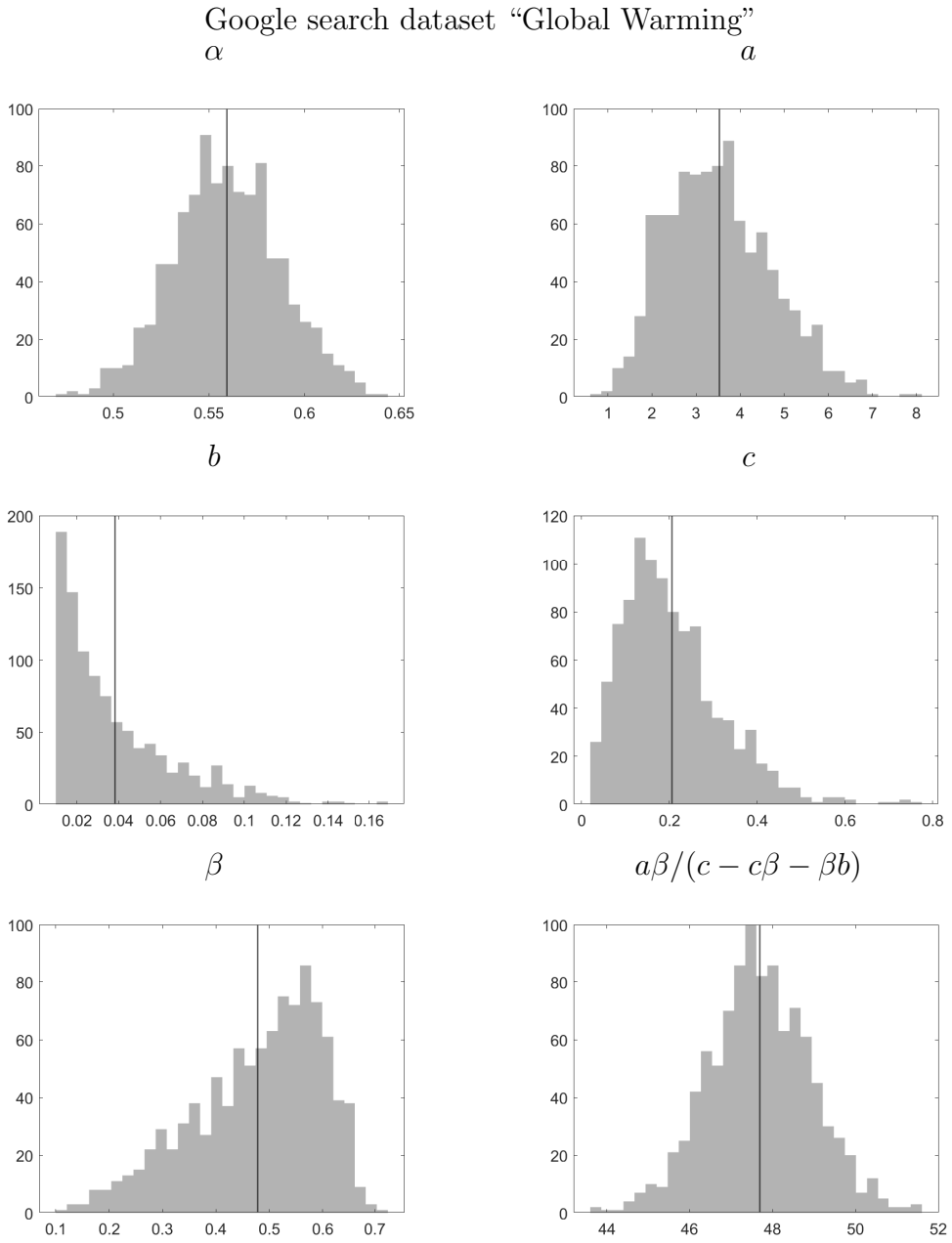


Figure 4.16: MCMC approximation of the posterior distribution (histogram) of the parameters. In all plots, the estimated value (vertical black solid).

Estratto per riassunto della tesi di dottorato

L'estratto (max. 1000 battute) deve essere redatto sia in lingua italiana che in lingua inglese e nella lingua straniera eventualmente indicata dal Collegio dei docenti.

L'estratto va firmato e rilegato come ultimo foglio della tesi.

Studente: Giulia Carallo

Matricola: 956387

Dottorato: Economia

Ciclo: 33° ciclo

Titolo della tesi: "Modelling Time Series of Counts"

Abstract: This thesis introduces new stochastic process with values in the set of integers. We introduce three different process to model the dynamics in the over-dispersion and under-dispersion feature in the data. In Chapter 1 we introduce an application of Generalized Poisson models for analysing over-dispersion in cyber-attacks. This chapter is motivational for the Generalised Poisson difference INGARCH model introduced in Chapter 2. In Chapter 3, we provide a generalisation defining a new Dynamic Conditional Score process with data distributed as Generalised Poisson. Finally, in Chapter 4 we introduce a new generalisation regarding the innovations. The model introduced is a Generalised Lagrangian Katz INAR process.

Abstract (italiano): La tesi introduce nuovi processi stocastici con valori nell'insieme dei numeri interi. Introduciamo tre diversi processi al fine di modellare l'eccessiva o sottodispersione presente nei dati. Nel Capitolo 1 introduciamo un applicazione del modello Generalized Poisson per l'analisi dell'eccessiva dispersione negli attacchi informatici. Questo capitolo è motivazionale per l'introduzione nel Capitolo 2 del modello Generalized Poisson Difference INGARCH. Nel Capitolo 3, forniamo una generalizzazione definendo un nuovo processo Dynamic Conditional Score con dati distribuiti come Generalised Poisson. Infine, nel Capitolo 4 introduciamo una nuova generalizzazione riguardo le innovazioni. Il modello introdotto è un processo Generalised Lagrangian Katz INAR.

Firma dello studente