



# Bayesian log-linear beta-negative binomial integer-valued Garch model

Yuanqi Chu<sup>1</sup> · Keming Yu<sup>1</sup>

Received: 4 May 2022 / Accepted: 27 June 2023  
© The Author(s) 2023

## Abstract

When dealing with time series with outlying and atypical data, a commonly used approach is to develop models based on heavy-tailed distributions. The literature coping with continuous-valued time series with extreme observations is well explored. However, current literature on modelling integer-valued time series data with heavy-tailedness is less considered. The state of the art research on this topic is presented by Gorgi (J R Stat Soc Ser B (Stat Methodol) 82:1325–1347, 2020) very recently, which introduced a linear Beta-negative binomial integer-valued generalized autoregressive conditional heteroscedastic (BNB-INGARCH) model. However, such proposed process allows for positive correlation only. This paper develops a log-linear version of the BNB-INGARCH model, which accommodates both negative and positive serial correlations. Moreover, we adopt Bayesian inference for better quantifying the uncertainty of unknown parameters. Due to the high computational demand, we resort to adaptive Markov chain Monte Carlo sampling schemes for parameter estimations and inferences. The performance of the proposed method is evaluated via a simulation study and empirical applications.

**Keywords** Beta-negative binomial distributions · Integer-valued GARCH models · Adaptive Markov chain Monte Carlo

## 1 Introduction

Despite a long history in the literature analyzing continuous time series variables, it is only in the recent years or so that much attention has been given to time series variables that are integer-valued (see Winkelmann 2008; Fahrmeir et al. 2013; Davis et al. 2016; Weiß 2018, and references therein for reviews).

---

✉ Keming Yu  
keming.yu@brunel.ac.uk

Yuanqi Chu  
yuanqi.chu@brunel.ac.uk

<sup>1</sup> Department of Mathematics, Brunel University London, Middlesex, UK

Generally, integer-valued time series models can be classified into two categories: “thinning” operator based models (Scotto et al. 2015) and regression based models (Fokianos 2012; Tjøstheim 2016).

To allow for dependence between time series data, two classes of models have been proposed in Cox et al. (1981): observation-driven models and parameter-driven models. In observation-driven models, the mean of the conditional distribution of the current observation  $y_t$  is directly specified as a function of past observations  $y_{t-1}, \dots, y_1$ . In parameter-driven models, dependence among observations is introduced via latent factors which follow a stochastic process, such as a hidden Markov chain (Leroux and Puterman 1992), or a latent stationary auto-regressive process (Zeger 1988; Chan and Ledolter 1995). Compared with parameter-driven models, observation-driven models are easier to fit in practical contexts with numerous covariates and long time series. See Zeger and Qaqish (1988) for a review of various observation-driven models for count time series data. A reference for the substantial development of observation-driven models can be found in Kedem and Fokianos (2005). A variety of observation-driven models for count responses have been developed. Davis et al. (2003, 2005) presented a class of observation-driven models for time series of Poisson counts and provided properties of the maximum likelihood estimators. Ahmad and Francq (2016) derived regularity conditions for the consistency and asymptotic normality (CAN) of the Poisson quasi-maximum likelihood estimator (QMLE) for time series of counts. However, the equidispersion assumption in Poisson distribution makes it too restrictive to be applied in empirical settings. Given this, Drescher (2005) considered various generalized count distributions for observation driven models and explored their maximum likelihood estimations. Regarding existing R packages (R Core Team 2021), the **glarma** package (Dunsmuir and Scott 2015) provides functions for estimation, testing, diagnostic checking and forecasting based on the generalized linear autoregressive moving average (GLARMA) class of observation-driven models for discrete-valued time series with regression variables.

The benchmark parameter-driven count data model introduced in Zeger (1988) has been widely extended. It has been considered as a class of the state-space model, which extends the generalized linear model by introducing a latent autoregressive process as the conditional mean function. The parameter-driven models allow the distribution of  $y_t$  to be dependent on this latent process and can deal with auto-correlation as well as over-dispersion in the model. However, parameter estimations in parameter-driven models require considerable computational effort. The main issue lies in the calculation requirement of very high dimensional integrals when using maximum likelihood estimation techniques, such that estimation methods based on Monte-Carlo (MC) integration are typically considered. To estimate the parameters of parameter-driven models, Chan and Ledolter (1995) employed a Monte Carlo EM algorithm. Kuk and Cheng (1997) considered the MC Newton Raphson method. However, such estimation approaches are not yet routinely available and therefore not ready for general use (Davis et al. 2003).

Bayesian estimation for time series of counts turns out to be a feasible and more elaborate alternative. Applications of Bayesian paradigm to count times series have mainly focused on parameter-driven models. Dynamic latent factor models within Bayesian count time series contexts have been actively studied (see, e.g., West and Harrison 1989; Durbin and Koopman 2000; Chib and Winkelmann 2001). Hay and Pettitt (2001) presented a fully Bayesian analysis of counts time series for a parameter-driven model with the form of a generalized linear mixed model, and investigated its application to the control of an infectious disease. Unlike the MC EM estimation approach for parameter-driven models, the Markov chain Monte Carlo (MCMC) procedure provides information of posterior distributions for both regression and time series parameters. In maximum likelihood-based estimations, estimation uncertainty is produced by constructing confidence intervals around the point forecasts. However, this kind of confidence intervals can only be justified asymptotically. When counts are small, such approximation is less accurate and the Bayesian technique arises as a prime candidate. When forecasting counts from the Bayesian perspective, not only the parameter uncertainty, but also the uncertainty caused by model specification, can be directly incorporated into the predictive probability mass function, which is a natural outcome of Bayes' theorem.

Although parameter-driven models are very flexible, the existence of unobservable latent factors brings a heavy computational burden even manageable via the MCMC sampling technique. On the other hand, Bayesian analysis of the more parsimonious observation-driven models has received growing attention recently. Generalized autoregressive moving average (GARMA) model extends the univariate Gaussian ARMA model to a flexible non-Gaussian observation-driven model. Silveira de Andrade et al. (2015) investigated the Bayesian approach for GARMA models with Poisson, binomial and negative binomial distributions. They utilized the Bayesian model selection criteria to choose the most appropriate model. Another advantage of Bayesian methodology over the corresponding frequentist procedure for forecasting discrete time series data is that Bayesian approach can produce only integer estimates of the count variable, while traditional forecasting often yields non-coherent (i.e. non-integer) estimates. For example, the Autoregressive Integrated Moving Average (ARIMA) model is one of the most prominent methods in financial time series forecasting. It has shown robust and efficient capability for short-term predictions and has been extensively applied to economics and finance fields (Contreras et al. 2003; Khashei et al. 2009; Lee and Ko 2011, among others). However, forecasts from ARIMA model can give negative values. Techniques such as log scale transformation or constrained forecast might guarantee non-negative predictions, but with the burden of elaborate post-processing and consequences of obtaining back-transformed forecasts that behave abnormally. Given the fact that many actual data in socioeconomic and business areas cannot have negative values, the classical ARIMA forecasting methods are improper when applied to non-negative count data series. From another point of view, Bayesians utilize the likelihood and prior multiplicity to generate forecasts from posterior predictive distributions by the iterative loop of MCMC procedures. Therefore, Bayesian model is prone to obey the non-negative value rules with its probabilistic predictive distributions, providing new perceptions for time series forecasting research. Nariswari and Pudjihastuti

(2019) implemented Bayesian time series estimation on ARIMA model for monthly medicine demand count data, showing the validity of Bayesian time series approach to avoid negative-value predictions, which is consistent with characteristics in the actual medicine data where the stock cannot have a negative value. McCabe and Martin (2005) developed Bayesian predictions of low count time series within the context of the integer-valued first-order autoregressive (INAR(1)) class of model, and showed the Bayesian method is feasible for producing coherent forecasts. Estimation uncertainty associated with both parameters and model specification is fully incorporated in their proposed methodology.

A commonly used model in most count time series data is the Poisson integer-valued generalized autoregressive conditional heteroscedastic (Poisson INGARCH) model proposed in Ferland et al. (2006). Since then, this model has been widely explored (see Neumann 2011; Doukhan et al. 2012, among others). However, the Poisson INGARCH model is not eligible to be applied in existence of potential extreme observations due to its equidispersion assumption. To this end, Zhu (2011) developed a negative binomial (NB) INGARCH model via the maximum likelihood approach. The NB-INGARCH model is flexible and allows for both overdispersion and extreme observations simultaneously. Later, Christou and Fokianos (2014) explored probabilistic properties and quasi-likelihood estimation for NB-INGARCH(1,1) process, and Xiong and Zhu (2019) considered a robust quasi-likelihood estimation for this process with an application to transaction counts. From a Bayesian perspective, Truong et al. (2017) proposed a hysteretic Poisson INGARCH model within the MCMC sampling scheme to estimate model parameters and adopted the Bayesian information criteria for model comparison. They highlighted their proposed model with a better performance of hysteresis in modelling the integer-valued time series. Chen et al. (2019) developed a Markov switching Poisson INGARCH model within a Bayesian framework to cope with the lagged dependence, overdispersion, consecutive zeros, non-linear dynamics and time varying coefficients for the meteorological variables. Some studies considered the natural candidates for the Poisson model. Chen and Lee (2017) proposed a Bayesian causality test based on the Poisson, negative binomial and log-linear Poisson INGARCH models with applications to climate and crime data. Recently, Chen and Khamthong (2020) introduced two nonlinear negative binomial INGARCH models (Markov switching and threshold specifications) along with the exogenous covariates in the conditional mean to describe time series of counts. They conducted parameter estimations and one-step-ahead forecasting via the Bayesian MCMC methods.

When modelling time series with outlying and atypical data, a commonly used approach is to develop models based on heavy-tailed distributions. The literature coping with continuous-valued time series with extreme observations is well explored via the Student's *t*-distribution (Harvey and Luati 2014). However, current literature on modelling discrete time series with heavy-tailedness is less considered. To fill this gap, very recently, Qian et al. (2020) proposed a new integer-valued autoregressive process with generalized Poisson-inverse Gaussian (GPiG) innovations to model heavy-tailed count time series data. Gorgi (2020) explored a heavy-tailed mixture of negative binomial distributions, known as the Beta-negative binomial (BNB) distribution, and developed a BNB auto-regression process for modelling integer-valued time series with

outliers, where a linear observation-driven dynamic equation for the conditional mean has been specified. Both Qian et al. (2020) and Gorgi (2020) employed maximum likelihood approaches to estimate the model parameters. This paper gives rise to the Bayesian inference for log-linear BNB-INGARCH models and conducts parameter estimations within adaptive Markov chain Monte Carlo frameworks. Moreover, the conditions for the posterior distribution of the full model parameter to be proper given some general priors have been derived and presented.

## 2 The log-linear BNB-INGARCH model

The Beta-negative binomial distribution can be represented as a beta mixture of the negative binomial distribution. Denote a discrete random variable  $Y$ , then  $Y \sim BNB(\beta, r, \alpha)$  if its probability mass function (PMF) is given by

$$f(Y = y) = \frac{\Gamma(y + r)}{\Gamma(y + 1)\Gamma(r)} \frac{B(\alpha + r, \beta + y)}{B(\alpha, \beta)}, y \in \mathcal{N} \tag{1}$$

where  $\Gamma(\cdot)$  is the gamma function and  $B(\cdot, \cdot)$  is the beta function.  $r > 0$  is the dispersion parameter,  $\alpha > 0$  is the tail parameter and  $\beta > 0$ . We follow the parameterization of the BNB distribution in terms of its mean parameter  $\lambda$  presented in Gorgi (2020)

$$f(Y = y) = \frac{\Gamma(y + r)}{\Gamma(y + 1)\Gamma(r)} \frac{B(\alpha + r, \frac{(\alpha-1)\lambda}{r} + y)}{B(\alpha, \frac{(\alpha-1)\lambda}{r})}, y \in \mathcal{N} \tag{2}$$

with mean  $\lambda > 0$ , dispersion  $r > 0$  and tail  $\alpha > 1$ . Let  $\{y_t\}, t = 1, \dots, n$  denote a univariate time series with the conditional distribution following the representation of  $BNB(\lambda_t, r, \alpha)$  at time  $t$ . We model the log-intensity process  $\mu_{t+1} = \log(\lambda_{t+1})$  in terms of a linear auto-regression process lying on its own past  $\mu_t$  and the past observation  $y_t$ . The log-linear BNB integer-valued generalized auto-regressive conditional heteroscedastic model of order (1,1) is defined by

$$\begin{aligned} y_t | \mathcal{F}_{t-1} &\sim BNB(\lambda_t, r, \alpha), \\ \mu_{t+1} &= \log(\lambda_{t+1}), \\ \mu_{t+1} &= \omega + \phi \log(y_t + 1) + \tau \mu_t \end{aligned} \tag{3}$$

where  $\mathcal{F}_{t-1}$  denotes the “ $\sigma$ -filed” generated by  $\{y_{t-1}, y_{t-2}, \dots\}$ . Here we follow Fokianos and Tjøstheim (2011) to choose  $\log(y_t + c)$  with constant  $c = 1$  in model (3) to map zeros of  $y_t$  into zeros of  $\log(y_t + 1)$ . Other reasonable choices for  $c$  may be considered. Note that model (3) accommodates both negative and positive serial correlations by allowing parameters  $\omega, \phi, \tau$  to take values in  $\mathbb{R}$ , whereas the linear BNB-INGARCH model introduced by Gorgi (2020) accommodates positive serial correlation only by restricting parameters to be positive to guarantee positivity of the conditional mean. Moreover, model (3) permits faster increase or decrease in

$\lambda_t$  according to the values of  $\omega, \phi$  and  $\tau$  than the linear model. Extensions to higher order log-linear BNB-INGARCH  $(p, q)$  models can be given as follows:

$$\begin{aligned}
 & y_t | \mathcal{F}_{t-1} \sim \text{BNB}(\lambda_t, r, \alpha), \\
 & \mu_{t+1} = \log(\lambda_{t+1}), \\
 & \mu_{t+1} = \omega + \sum_{i=1}^p \phi_i \log(y_{t+1-i} + 1) + \sum_{j=1}^q \tau_j \mu_{t+1-j}
 \end{aligned} \tag{4}$$

We note that the BNB distribution belongs to the class of mixed Poisson distributions and can approximate arbitrarily well the negative binomial distribution as well as the Poisson distribution. Specifically, as  $\alpha \rightarrow \infty$ , the parameterized distribution  $\text{BNB}(\lambda, r, \alpha)$  converges to a NB distribution with dispersion  $r$  and success probability  $\lambda/(\lambda + r)$ . Furthermore, as  $r \rightarrow \infty$ , the BNB converges to a Poisson distribution with mean  $\lambda$  (Gorgi 2020). Given this Poisson approximation to the BNB distribution, we follow (Douc et al. (2013), Lemma 14) and Liboschik et al. (2017) to impose the conditions  $\{|\phi|, |\tau| < 1, |\phi + \tau| < 1\}$  and  $\{|\phi_i|, |\tau_j| < 1, |\sum_{i=1}^p \phi_i + \sum_{j=1}^q \tau_j| < 1\}$  to guarantee stationarity of the proposed processes (3) and (4) respectively. We further follow Wang et al. (2014) and Gorgi (2020) to preassign the initial point  $\lambda_1$  to a fixed positive value for both models (3) and (4). As noted in Gorgi (2020), this approach is quite standard in the literature of observation-driven time series models. In fact, Gorgi (2020) showed that the filtered parameter  $\{\hat{\lambda}_t(\theta)\}_{t \in \mathbb{N}}$  converges exponentially almost surely and uniformly over the compact parameter sets  $\Theta$  to a unique stationary and ergodic sequence  $\{\tilde{\lambda}_t(\theta)\}_{t \in \mathbb{Z}}$  for any initialization  $\hat{\lambda}_1(\theta) \in \mathbb{R}^+$ .

### 3 Bayesian inference

Due to the high computational demand, we resort to Bayesian analysis for parameter estimations and inferences of the log-linear BNB-INGARCH processes. Without loss of generality, we focus on the first-order specification  $(p = q = 1)$  as presented in model (3) for simplicity of inference illustration. We denote the time series of interest as  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  and  $\theta = (\omega, \phi, \tau, r, \alpha)^T$  as the entire unknown parameter vector. By the Bayes theorem, the posterior distribution  $p(\theta|\mathbf{y})$  is given by

$$p(\theta|\mathbf{y}) \propto L(\mathbf{y}|\theta)\pi(\theta) \tag{5}$$

where  $\pi(\theta)$  denotes the prior distribution and  $L(\mathbf{y}|\theta)$  represents the likelihood function

$$L(\mathbf{y}|\theta) = \prod_{t=2}^n \frac{\Gamma(y_t + r)}{\Gamma(r)} \frac{B\left(\alpha + r, \frac{(\alpha-1)e^{\omega+\phi \log(y_{t-1}+1)+\tau \log(\lambda_{t-1})}}{r} + y_t\right)}{B\left(\alpha, \frac{(\alpha-1)e^{\omega+\phi \log(y_{t-1}+1)+\tau \log(\lambda_{t-1})}}{r}\right)} \tag{6}$$

with  $\omega \in \mathbb{R}$ ,  $|\phi| < 1$ ,  $|\tau| < 1$ ,  $|\phi + \tau| < 1$ ,  $r > 0$  and  $\alpha > 1$ . We implement Bayesian inference for parameter groups (i)  $\omega$ ; (ii)  $\{\phi, \tau\}$ ; (iii)  $r$ ; and (iv)  $\alpha$  with the assumption that they are priori-independent. The conditional posterior distributions for each parameter group are presented as follows. Hereafter let  $\theta_j$  denote the  $j$ -th element of the parameter vector  $\theta$ , with  $j = 1, 2, 3, 4$  referring to  $\omega$ ,  $\{\phi, \tau\}$ ,  $r$  and  $\alpha$  respectively, and  $\theta_{-j}$  denote the vector of all parameters excluding the component  $\theta_j$ .

- (i) For the intercept  $\omega$ , we consider a normal prior  $\pi(\omega) = N(\mu_\omega, \sigma_\omega^2)$ . The full conditional posterior distribution of  $\omega$  is then as follows:

$$\begin{aligned}
 p(\omega|\theta_{-1}, \mathbf{y}) &\propto L(\mathbf{y}|\theta)\pi(\omega) \\
 &\propto L(\mathbf{y}|\theta)e^{-\frac{(\omega-\mu_\omega)^2}{2\sigma_\omega^2}}
 \end{aligned}
 \tag{7}$$

- (ii) For the prior  $\pi(\phi, \tau)$  of the parameter block  $\{\phi, \tau\}$ , we employ constrained normals  $\pi(\phi) = N(\mu_\phi, \sigma_\phi^2)$  and  $\pi(\tau) = N(\mu_\tau, \sigma_\tau^2)$  with  $\{\phi, \tau\}$  satisfying the set  $A$

$$|\phi| < 1, |\tau| < 1, |\phi + \tau| < 1 \tag{8}$$

Then the full conditional posterior distribution of  $\{\phi, \tau\}$  is given by

$$\begin{aligned}
 p(\phi, \tau|\theta_{-2}, \mathbf{y}) &\propto L(\mathbf{y}|\theta)\pi(\phi, \tau) \\
 &\propto L(\mathbf{y}|\theta)e^{-\frac{(\phi-\mu_\phi)^2}{2\sigma_\phi^2}} e^{-\frac{(\tau-\mu_\tau)^2}{2\sigma_\tau^2}} \mathcal{I}(A)
 \end{aligned}
 \tag{9}$$

where  $\mathcal{I}(\cdot)$  denotes the indicator function.

- (iii) For the dispersion parameter  $r$ , we impose a gamma prior  $\pi(r) = \text{Ga}(\eta_{1r}, \eta_{2r})$  with shape parameter  $\eta_{1r}$  and rate parameter  $\eta_{2r}$ , then the full conditional posterior distribution of  $r$  is given by

$$\begin{aligned}
 p(r|\theta_{-3}, \mathbf{y}) &\propto L(\mathbf{y}|\theta)\pi(r) \\
 &\propto L(\mathbf{y}|\theta)r^{\eta_{1r}-1} e^{-\eta_{2r} \cdot r}
 \end{aligned}
 \tag{10}$$

- (iv) For the tail parameter  $\alpha$ , we impose a truncated gamma prior  $\pi(\alpha) = \text{Ga}(\eta_{1\alpha}, \eta_{2\alpha})\mathcal{I}(\alpha > 1)$  with shape  $\eta_{1\alpha}$  and rate  $\eta_{2\alpha}$ , then the full conditional posterior distribution of  $\alpha$  is given by

$$\begin{aligned}
 p(\alpha|\theta_{-4}, \mathbf{y}) &\propto L(\mathbf{y}|\theta)\pi(\alpha) \\
 &\propto L(\mathbf{y}|\theta)\alpha^{\eta_{1\alpha}-1} e^{-\eta_{2\alpha} \cdot \alpha} \mathcal{I}(\alpha > 1)
 \end{aligned}
 \tag{11}$$

Theorem 1 elaborated below presents the sufficient conditions for the posterior distribution of  $\theta$  to be proper given some general priors.

**Theorem 1** *Let  $\{y_t|\mathcal{F}_{t-1}\}_{t \in \mathbb{Z}^+}$  denote the target count time series with the conditional Beta-negative binomial distribution and  $\theta = (\omega, \phi, \tau, r, \alpha)^T$  be the full parameter vector. For ease of notation, we denote  $\kappa(\theta) = (\alpha - 1)/r \cdot e^{\omega+\phi \log(y_{t-1}+1)+\tau \log(\lambda_{t-1})}$  in*

the remainder of this section. Then under model (3) and proper prior specifications, the posterior distribution of  $\theta$  given  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  is obtained by

$$p(\theta|\mathbf{y}) \propto \prod_{i=2}^n \frac{\Gamma(y_i + r)}{\Gamma(r)} \frac{B(\alpha + r, \kappa(\theta) + y_i)}{B(\alpha, \kappa(\theta))} \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r)$$

and is well defined.

**Proof** Under proper prior specifications, we have

$$\begin{aligned} & \int p(\theta|\mathbf{y})d\theta \\ & \propto \int \left\{ \prod_{i=2}^n \frac{\Gamma(y_i + r)}{\Gamma(r)} \frac{B(\alpha + r, \kappa(\theta) + y_i)}{B(\alpha, \kappa(\theta))} \times \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) \right\} d\omega d\phi d\tau d\alpha dr \\ & = \int \left\{ \prod_{i=2}^n \frac{\Gamma(y_i + r)}{\Gamma(r)} \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \frac{\Gamma(\kappa(\theta) + y_i)}{\Gamma(\kappa(\theta))} \frac{\Gamma(\alpha + \kappa(\theta))}{\Gamma(\alpha + \kappa(\theta) + y_i + r)} \pi(\omega)\pi(\phi, \tau) \right. \\ & \left. \pi(\alpha)\pi(r) \right\} d\omega d\phi d\tau d\alpha dr \end{aligned} \tag{12}$$

where the equation follows by the relation  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ . From the Stirling’s approximation of the gamma function  $\Gamma(\cdot)$ , the following approximation is obtained for large  $r$  and  $\alpha$ :

$$\begin{aligned} (12) & \approx \int r^{\sum_{i=2}^n y_i} \left\{ \prod_{i=2}^n \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \frac{\Gamma(\kappa(\theta) + y_i)}{\Gamma(\kappa(\theta))} \frac{\Gamma(\alpha + \kappa(\theta))}{\Gamma(\alpha + \kappa(\theta) + y_i + r)} \right. \\ & \left. \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) \right\} d\omega d\phi d\tau d\alpha dr \\ & \approx \int r^{\sum_{i=2}^n y_i} \left\{ \prod_{i=2}^n \alpha^r (\kappa(\theta))^{y_i} (\alpha + \kappa(\theta))^{-(y_i+r)} \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) \right\} \\ & d\omega d\phi d\tau d\alpha dr \\ & = \int r^{\sum_{i=2}^n y_i} \left\{ \prod_{i=2}^n \left( \frac{\alpha}{\alpha + \kappa(\theta)} \right)^r \left( \frac{\kappa(\theta)}{\alpha + \kappa(\theta)} \right)^{y_i} \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) \right\} \\ & d\omega d\phi d\tau d\alpha dr \\ & = \int r^{\sum_{i=2}^n y_i} \left\{ \prod_{i=2}^n \left( \frac{1}{1 + \frac{\kappa(\theta)}{\alpha}} \right)^r \left( \frac{1}{1 + \frac{\alpha}{\kappa(\theta)}} \right)^{y_i} \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) \right\} \\ & d\omega d\phi d\tau d\alpha dr \end{aligned} \tag{13}$$

For  $r \geq 1$ , we obtain



$$\begin{aligned}
 (13) &\leq \int r^{\sum_{i=2}^n y_i} \left\{ \prod_{i=2}^n \frac{1}{1 + \frac{r\kappa(\theta)}{\alpha}} \left( \frac{1}{1 + \frac{\alpha}{\kappa(\theta)}} \right)^{y_i} \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) \right\} \\
 &\quad d\omega d\phi d\tau d\alpha dr \\
 &\leq \int r^{\sum_{i=2}^n y_i} \left\{ \prod_{i=2}^n \frac{1}{1 + \frac{r\kappa(\theta)}{\alpha}} \frac{1}{1 + \frac{\alpha y_i}{\kappa(\theta)}} \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) \right\} \\
 &\quad d\omega d\phi d\tau d\alpha dr
 \end{aligned}
 \tag{14}$$

where the first inequality follows by Bernoulli's inequality  $(1 + x)^d \geq 1 + dx$  for real numbers  $d \geq 1, x \geq -1$  and the second inequality follows by Bernoulli's inequality  $(1 + x)^d \geq 1 + dx$  for integer  $d \geq 0$  and real number  $x \geq -1$ . Without loss of generality, we assume that the first  $n_1$   $y_i$ 's have zero-valued observations and the remaining  $(n - n_1)$  observations have positive  $y_i$ 's. Then we have

$$\begin{aligned}
 (14) &= \int r^{\sum_{i=n_1+1}^n y_i} \left\{ \prod_{i=2}^{n_1} \frac{1}{1 + \left(\frac{\alpha-1}{\alpha}\right) e^{\omega+\tau \log(\lambda_{i-1})}} \prod_{i=n_1+1}^n \frac{1}{1 + \frac{r\kappa(\theta)}{\alpha}} \frac{1}{1 + \frac{\alpha y_i}{\kappa(\theta)}} \right. \\
 &\quad \left. \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) \right\} d\omega d\phi d\tau d\alpha dr \\
 &\leq \int r^{\sum_{i=n_1+1}^n y_i} \left\{ \prod_{i=2}^{n_1} \frac{1}{\left(\frac{\alpha-1}{\alpha}\right) e^{\omega+\tau \log(\lambda_{i-1})}} \prod_{i=n_1+1}^n \frac{\alpha}{r\kappa(\theta)} \frac{\kappa(\theta)}{\alpha y_i} \pi(\omega)\pi(\phi, \tau) \right. \\
 &\quad \left. \pi(\alpha)\pi(r) \right\} d\omega d\phi d\tau d\alpha dr \\
 &= \int r^{\sum_{i=n_1+1}^n y_i} \left\{ \prod_{i=2}^{n_1} \frac{\alpha e^{-\{\omega+\tau \log(\lambda_{i-1})\}}}{\alpha - 1} \prod_{i=n_1+1}^n \frac{\alpha e^{-\{\omega+\phi \log(y_{i-1}+1)+\tau \log(\lambda_{i-1})\}}}{\alpha - 1} \right. \\
 &\quad \left. \frac{(\alpha - 1) e^{\omega+\phi \log(y_{i-1}+1)+\tau \log(\lambda_{i-1})}}{\alpha r y_i} \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) \right\} d\omega d\phi d\tau d\alpha dr \\
 &= C_1 \int r^{\sum_{i=n_1+1}^n (y_i-1)} \left( \frac{\alpha}{\alpha - 1} \right)^{n_1-1} e^{-(n_1-1)\omega} e^{-(\sum_{i=2}^{n_1} \log(\lambda_{i-1}))\tau} \\
 &\quad \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) d\omega d\phi d\tau d\alpha dr
 \end{aligned}
 \tag{15}$$

where  $C_1 = \prod_{i=n_1+1}^n \frac{1}{y_i}$ . By the Binomial approximation  $\left(\frac{\alpha}{\alpha-1}\right)^{n_1-1} = \left(1 + \frac{1}{\alpha-1}\right)^{n_1-1} \approx 1 + \frac{n_1-1}{\alpha-1}$  for large  $\alpha$  and further assuming  $\alpha \geq M$  for any large positive number  $M$ , we obtain

$$\begin{aligned}
 (15) &\approx C_1 \int r^{\sum_{t=n_1+1}^n (y_t-1)} \left(1 + \frac{n_1-1}{\alpha-1}\right) e^{-(n_1-1)\omega} e^{-(\sum_{t=2}^{n_1} \log(\lambda_{t-1}))\tau} \\
 &\quad \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) d\omega d\phi d\tau d\alpha dr \\
 &\leq C_1 \int r^{\sum_{t=n_1+1}^n (y_t-1)} \left(1 + \frac{n_1-1}{M-1}\right) e^{-(n_1-1)\omega} e^{-(\sum_{t=2}^{n_1} \log(\lambda_{t-1}))\tau} \\
 &\quad \pi(\omega)\pi(\phi, \tau)\pi(\alpha)\pi(r) d\omega d\phi d\tau d\alpha dr \\
 &= C_2 \int r^{\sum_{t=n_1+1}^n (y_t-1)} e^{-(n_1-1)\omega} \pi(\omega)\pi(r) d\omega dr < \infty
 \end{aligned}$$

given that the prior density functions  $\pi(\omega)$ ,  $\pi(\phi, \tau)$ ,  $\pi(\alpha)$  and  $\pi(r)$  integrate to a finite quantity, the  $\sum_{t=n_1+1}^n (y_t - 1)$ -th moment about the origin of  $r$  exists and the moment generating function  $M_\omega(1 - n_1)$  of  $\omega$  exists. Here  $C_2$  is a constant unrelated with the parameters of interest.

It follows from the proof of Theorem 1 that the priors for  $(\phi, \tau)$  and  $\alpha$  can be chosen very flexibly. In fact, any proper distributions can be considered because the appropriateness of the posterior will not be affected. On the other hand, the choice of the prior  $\pi(r)$  requires existence of the  $\sum_{t=n_1+1}^n (y_t - 1)$ -th moment about the origin of  $r$  and the choice of the prior  $\pi(\omega)$  requires existence of the moment generating function  $M_\omega(1 - n_1)$ .

Since the obtained posterior distributions do not correspond to closed-form distributions, we resort to MCMC sampling methods. Specifically, for faster convergence and better mixing of the chain, we follow the adaptive MCMC method of Chen and So (2006). We employ the random-walk Metropolis-Hastings in the first  $H$  iterations (the burn-in period) and the independent-kernel Metropolis-Hastings in the following  $(N - H)$  iterations to draw samples for the parameter vector  $\theta$ . The adaptive MCMC procedure is provided as follows:

- Step1. Set initial values for  $\theta^{(0)} = (\omega^{(0)}, \phi^{(0)}, \tau^{(0)}, r^{(0)}, \alpha^{(0)})^T$ .
- Step2. When  $1 \leq k \leq H$ , we adopt the random-walk Metropolis-Hastings algorithm for sampling  $\theta^{(k)}$ :
  - Step2.1. Generate candidate values  $\theta^* = \{\theta_l^*\}$ ,  $l = 1, 2, \dots, 5$ , where  $\theta_l^* = \theta_l^{(k-1)} + \epsilon$ ,  $\epsilon \sim N(0, \sigma_l^2)$ , and the tuning parameter  $\sigma_l^2$  is selected to achieve the acceptance rate around 23%.
  - Step2.2. Keep the candidate values if  $\theta^*$  satisfies that:  $\theta_1 > 0$ ,  $\{\theta_2, \theta_3\}$  do not violate the stationarity conditions of the model,  $\theta_4 > 0$  and  $\theta_5 > 1$ . Otherwise, go back to Step 2.1.
  - Step2.3. Calculate the acceptance probability

$$\text{Prob}(\theta^*, \theta^{(k-1)}) = \min \left\{ 1, \frac{p(\theta^*)}{p(\theta^{(k-1)})} \right\},$$

where  $p(\cdot)$  is the target posterior distribution given in (5). Then generate a random uniform number  $u \in [0, 1]$ . If  $u < \text{Prob}(\theta^*, \theta^{(k-1)})$ , accept the new candidate and set  $\theta^{(k)} = \theta^*$ . Otherwise, set  $\theta^{(k)} = \theta^{(k-1)}$ .

Step3. When  $k \geq H + 1$ , we adopt the independent-kernel Metropolis-Hastings algorithm for sampling  $\theta^{(k)}$ :

Step3.1. Generate candidate values  $\theta^* \sim N(\mu_\theta, \Omega_\theta)$ , where the sample mean  $\mu_\theta$  and sample covariance matrix  $\Omega_\theta$  are calculated using the burn-in  $H$  iteration samples.

Step3.2. Keep the candidate values if  $\theta^*$  satisfies that:  $\theta_1 > 0$ ,  $\{\theta_2, \theta_3\}$  do not violate the stationarity conditions of the model,  $\theta_4 > 0$  and  $\theta_5 > 1$ . Otherwise, go back to Step 3.1.

Step3.3. Calculate the acceptance probability

$$\text{Prob}(\theta^*, \theta^{(k-1)}) = \min \left\{ 1, \frac{p(\theta^*)g(\theta^{(k-1)})}{p(\theta^{(k-1)})g(\theta^*)} \right\},$$

where  $g(\cdot)$  is the Gaussian proposal density with mean  $\mu_\theta$  and sample covariance matrix  $\Omega_\theta$ . Then generate a random uniform number  $u \in [0, 1]$ . If  $u < \text{Prob}(\theta^*, \theta^{(k-1)})$ , accept the new candidate and set  $\theta^{(k)} = \theta^*$ . Otherwise, set  $\theta^{(k)} = \theta^{(k-1)}$ .

Step4. Go to the next iteration or stop if the chain has converged.

Given the obtained  $N$  iteration samples  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}$ , we discard the first  $H$  in the burn-in period and perform model parameter estimations using the remaining  $(N - H)$  iterations.

#### 4 Simulation analysis

To examine the performance of the adaptive MCMC algorithm for the log-linear BNB-INGARCH model, we conduct a simulation analysis under the following data generating process (DGP) with sample sizes  $n = 100$  and  $n = 250$ . The count series  $y_t$  is sampled from the log-linear BNB-INGARCH model (3) where  $(\omega, \phi, \tau, r, \alpha)^T$  is set to be  $(0.65, 0.7, -0.2, 5, 3)^T$ . To investigate the sensitivity of the prior and

hyperparameter selections, we consider three prior calibrations for the parameters of interest, which is:

Prior 1:  $\omega \sim N(0.1, 0.3^2)$ ,  $\{\phi, \tau\} \sim N(0.1, 0.25^2) \cdot N(0.1, 0.75^2)\mathcal{I}(A)$ ,  
 $r \sim \text{Ga}(10, 0.5)$  and  $\alpha \sim \text{Ga}(10, 0.5)\mathcal{I}(\alpha > 1)$

Prior 2:  $\omega \sim N(0.1, 0.3^2)$ ,  $\pi(\phi, \tau) \propto \mathcal{I}(A)$ ,  $r \sim \text{Ga}(10, 0.5)$  and  
 $\alpha \sim \text{Exp}(0.01)\mathcal{I}(\alpha > 1)$

Prior 3:  $\omega \sim N(0.1, 0.3^2)$ ,  $\{\phi, \tau\} \sim N(0.1, 0.25^2) \cdot N(0.1, 0.75^2)\mathcal{I}(A)$ ,  
 $r \sim \text{Ga}(10, 1)$  and  $\pi(\alpha) \propto \frac{1}{(1+\alpha)^2}\mathcal{I}(\alpha > 1)$

where  $\text{Exp}(\eta)$  denotes the exponential distribution with the rate parameter  $\eta$  and the set  $A$  is given in (8). Note that in Prior 2, we employ a constrained uniform prior on  $\{\phi, \tau\}$  which configures a flat prior on the parameters restricted by the indicator  $\mathcal{I}(A)$ . We perform 10,000 MCMC iterations by discarding 5000 iterations as a burn-in sample for inference in each scenario. Table 1 presents the averages of posterior means, medians, standard deviations, and 95% credible intervals over 1000 replications for all model parameters. For a comparison with the frequentist approach, we report the simulation results for the maximum likelihood (ML) estimator of the considered model in Table 2. The averages of the mean and the standard deviation for different parameter values and sample sizes are obtained from 1000 Monte Carlo replications. We follow Gorgi (2020) to reparameterize  $r$  and  $\alpha$  in terms of their inverse to facilitate the ML estimation. As noted by Gorgi (2020), this procedure is adopted because, especially for small sample sizes, the estimate of  $\alpha$  may become arbitrarily large since the likelihood function is flat for large  $\alpha$ .

Given Tables 1 and 2, we summarize the simulation results as follows:

(1) We observe that the employed adaptive MCMC approach gives a reasonably accurate estimate of all the parameters of interest for small sample sizes. The standard deviation decreases as the sample size increases from 100 to 250 in both Bayesian and ML estimation scenarios. Furthermore, all the reported standard deviation values for parameters  $\omega$ ,  $\phi$  and  $\tau$  under Bayesian estimation are smaller than that of ML estimator in each simulation scheme.

(2) The simulation results demonstrate that both positive and negative serial correlations can be captured by the proposed model. Moreover, the average posterior means and posterior medians are overall reasonably close to the true values of the parameters, implying the validity of the considered adaptive MCMC method. We therefore suggest using posterior medians as the model parameter estimates since the median is a robust measure of central tendency compared with the mean.

(3) We observe that the adaptive MCMC procedure is robust to the selection of priors and hyperparameters via delivering similar reasonably accurate estimation results under different setting scenarios.

(4) We extensively examine the sensitivity of starting values by specifying different starting points for the adaptive MCMC sampler. We also investigate the sensitivity relating to the choice of  $\lambda_1$  by randomly setting different initialization values for the intensity process. We observe that the employed MCMC sampler is robust to the selection of the starting values and the initial intensity as different calibrations deliver similar reasonably accurate estimation results for the considered small

**Table 1** Simulation results for the Bayesian log-linear BNB-INGARCH model obtained from 1000 replications

Parameter	True value	Ave mean	Ave median	Ave Std	95% CI		Ave Mean	Ave Median	Ave Std	95% CI	
					P2.5	P97.5				P2.5	P97.5
$n = 100$							$n = 250$				
	Prior 1						Prior 1				
$\omega$	0.65	0.6378	0.6310	0.1587	0.3576	0.9430	$\omega$	0.6630	0.1326	0.4320	0.9234
$\phi$	0.7	0.7215	0.7257	0.1095	0.5129	0.8982	$\phi$	0.7212	0.0791	0.5849	0.8695
$\tau$	-0.2	-0.1013	-0.1026	0.1415	-0.3421	0.1619	$\tau$	-0.1824	0.1033	-0.3654	0.0088
$r$	5	4.4248	4.4779	0.5285	3.3239	5.1603	$r$	4.1335	0.4844	3.1148	4.8305
$\alpha$	3	2.5263	2.5369	0.1192	2.2712	2.6833	$\alpha$	2.4937	0.0974	2.2695	2.6132
$n = 100$							$n = 250$				
	Prior 2						Prior 2				
$\omega$	0.65	0.6262	0.6356	0.1741	0.3135	0.9431	$\omega$	0.6081	0.1468	0.3568	0.8770
$\phi$	0.7	0.8709	0.8847	0.0801	0.7083	0.9826	$\phi$	0.8173	0.0770	0.6739	0.9517
$\tau$	-0.2	-0.2595	-0.2667	0.1216	-0.4598	-0.0188	$\tau$	-0.2595	0.0982	-0.4329	-0.0750
$r$	5	5.3240	5.4157	0.7225	3.7592	6.3074	$r$	4.4359	0.6042	3.1873	5.2949
$\alpha$	3	2.7084	2.7409	0.1659	2.3140	2.8927	$\alpha$	2.6120	0.1359	2.3061	2.7877
$n = 100$							$n = 250$				
	Prior 3						Prior 3				
$\omega$	0.65	0.6133	0.6081	0.1841	0.2898	0.9537	$\omega$	0.6294	0.1441	0.3743	0.9170
$\phi$	0.7	0.7372	0.7444	0.1146	0.5151	0.9372	$\phi$	0.7320	0.0785	0.5933	0.8788
$\tau$	-0.2	-0.1452	-0.1622	0.1555	-0.4058	0.1589	$\tau$	-0.2122	0.1074	-0.4116	-0.0189
$r$	5	4.8441	4.9470	0.7813	3.2057	5.9008	$r$	4.2510	0.6038	2.9771	5.1067
$\alpha$	3	2.8488	2.8851	0.2583	2.3006	3.2031	$\alpha$	2.7703	0.2056	2.3320	3.0563

Prior 1:  $\omega \sim N(0.1, 0.3^2)$ ,  $\{\phi, \tau\} \sim N(0.1, 0.25^2) \cdot N(0.1, 0.75^2) \mathcal{I}(A)$ ,  $r \sim \text{Ga}(10, 0.5)$  and  $\alpha \sim \text{Ga}(10, 0.5) \mathcal{I}(\alpha > 1)$

Prior 2:  $\omega \sim N(0.1, 0.3^2)$ ,  $\pi(\phi, \tau) \propto \mathcal{I}(A)$ ,  $r \sim \text{Ga}(10, 0.5)$  and  $\alpha \sim \text{Exp}(0.01) \mathcal{I}(\alpha > 1)$

Prior 3:  $\omega \sim N(0.1, 0.3^2)$ ,  $\{\phi, \tau\} \sim N(0.1, 0.25^2) \cdot N(0.1, 0.75^2) \mathcal{I}(A)$ ,  $r \sim \text{Ga}(10, 1)$  and  $\pi(\alpha) \propto \frac{1}{(1+\alpha)^2} \mathcal{I}(\alpha > 1)$

**Table 2** Simulation results for the ML estimators of the log-linear BNB-INGARCH model obtained from 1000 Monte Carlo replications. The parameters  $r$  and  $\alpha$  are reparameterized in terms of their inverse

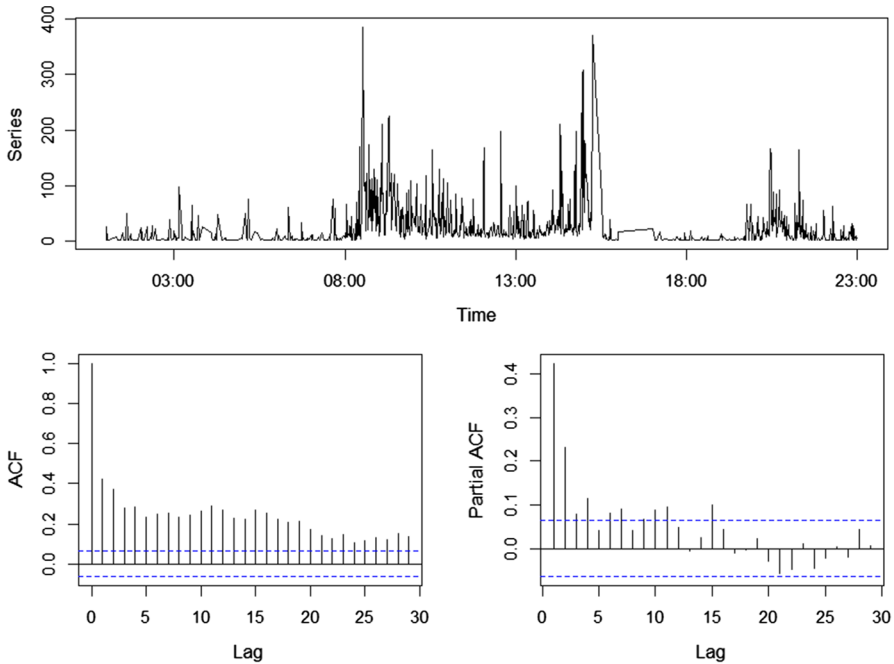
Parameter	True Value	Ave Mean	Ave Std	True Value	Ave Mean	Ave Std	
$n = 100$				$n = 250$			
$\omega$	0.65	0.6407	0.2976	$\omega$	0.65	0.6412	0.1846
$\phi$	0.7	0.6945	0.1286	$\phi$	0.7	0.6914	0.0874
$\tau$	-0.2	-0.2392	0.1830	$\tau$	-0.2	-0.2107	0.1075
$r^{-1}$	0.2	0.1692	0.2740	$r^{-1}$	0.2	0.1613	0.2001
$\alpha^{-1}$	0.3333	0.1454	0.1552	$\alpha^{-1}$	0.3333	0.1992	0.1379

sample sizes  $n$ . For space limitation, we do not present these results, but they are available from the authors upon request.

## 5 Empirical application on futures tick count data

This section illustrates the proposed methodology by an empirical application on the numbers of minute-bar VIX futures tick count data. This historical intraday market data is delivered by Tick Data provider and is available from <https://www.tickdata.com/>. The sample data set we consider consists of 920 available observations between 01:00 a.m. and 11:00 p.m. on the day 02nd January, 2020. The empirical mean and variance are 25.789 and 1669.625 respectively, indicating considerable over-dispersion pattern of the data set. Figure 1 depicts the plot and the empirical auto-correlation functions of the tick count series. The auto-correlation function (ACF) and partial auto-correlation function (PACF) plots suggest the existence of significant auto-correlations in the data. The series exhibits several extreme observations. Specifically, the number of minute-bar VIX futures tick count is exceedingly high at 08:31 a.m., 14:58 p.m. and 15:15 p.m.. These attributes indicate the desirability of BNB auto-regressions to capture the auto-correlation structure and to account for the extreme observations in the data by means of the heavy-tailedness characteristics of the BNB distribution.

We compare the performances of Bayesian log-linear BNB-INGARCH models and their counterpart log-linear NB-INGARCH models with order specifications  $\{(p, q)\} = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$ . The prior calibrations are presented below. For both BNB-INGARCH and NB-INGARCH models, we adopt normal prior  $N(0.5, 0.25^2)$  for  $\omega$ , constrained normal priors  $N(0.2, 0.15^2)$  for  $\phi_i$  and  $\tau_j$  with  $\{\phi_i, \tau_j\}$  satisfying  $|\phi_i| < 1, |\tau_j| < 1, |\sum_{i=1}^p \phi_i + \sum_{j=1}^q \tau_j| < 1$  and gamma prior  $\text{Ga}(5, 0.5)$  for  $r$ . For the tail parameter  $\alpha$  under BNB model specifications, we consider the truncated gamma prior  $\text{Ga}(5, 0.5)\mathcal{I}(\alpha > 1)$ . We also refer to the analogical prior establishments in the simulation analysis section for the considered empirical application and observe the parameter estimations are robust to the selection of priors and hyperparameters. We perform 30,000 MCMC iterations and discard the first 10,000 iterations as a burn-in sample in each model set-up scenario. Table 3 summarizes the estimation results including the posterior mean, standard deviation, the posterior



**Fig. 1** Numbers of minute-bar VIX futures tick count data on 02nd January, 2020 from 01:00 a.m. to 11:00 p.m. (top panel) and sample auto-correlation function (bottom left) and partial auto-correlation function of the series (bottom right)

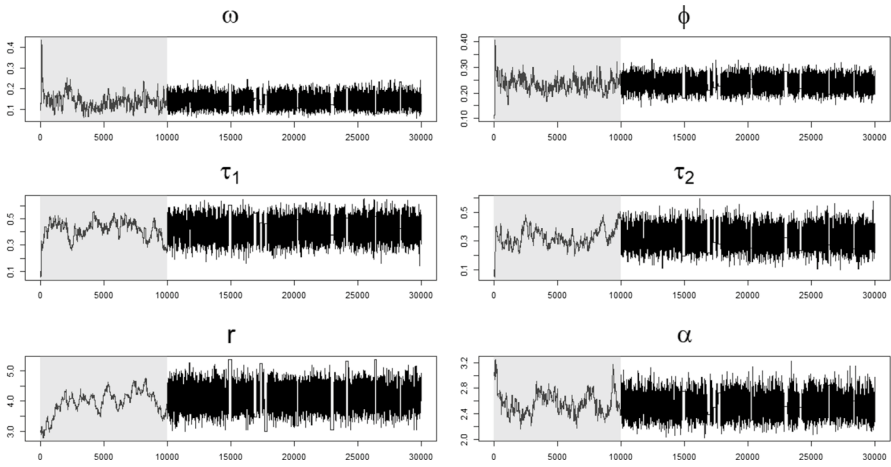
2.5 and 97.5 percentiles, Akaike information criterion (AIC) and Bayesian information criterion (BIC) for each specification separately. The convergence diagnostics of each of the Markov chains is investigated through the trace plots for each parameter of the fitted models. We conclude that proper mixing is achieved by each MCMC sampler. Based on the values of AIC and BIC provided in Table 3, the BNB-INGARCH models under all considered order specifications are superior to their corresponding NB-INGARCH counterparts. The preference for the BNB distribution can be further embodied in the low estimate values of the tail parameter  $\alpha$ , which is estimated to be around 2.5 with a standard deviation of about 0.15 in most order specification scenarios. This implies the heavy-tailedness of the estimated conditional distribution of the data with only a finite second order moment. Moreover, the BNB-INGARCH(1,2) is found to be a competitive model for this data set among all the considered candidates. Due to space limitation, we only provide the diagnostic trace plots of the Markov chains for each parameter in the favoured BNB-INGARCH(1,2) model in Fig. 2.

To check the adequacy of the best fitted log-linear BNB-INGARCH(1,2) model based on the lowest AIC and BIC values, we examine the standardized Pearson residuals  $z_t = \frac{y_t - E(y_t | \mathcal{F}_{t-1})}{\sqrt{\text{Var}(y_t | \mathcal{F}_{t-1})}}$  proposed by Jung et al. (2006). For correctly specified model, the Pearson residuals should have mean zero and variance one, with no significant auto-correlations. Figure 3 presents the series plot and auto-correlation

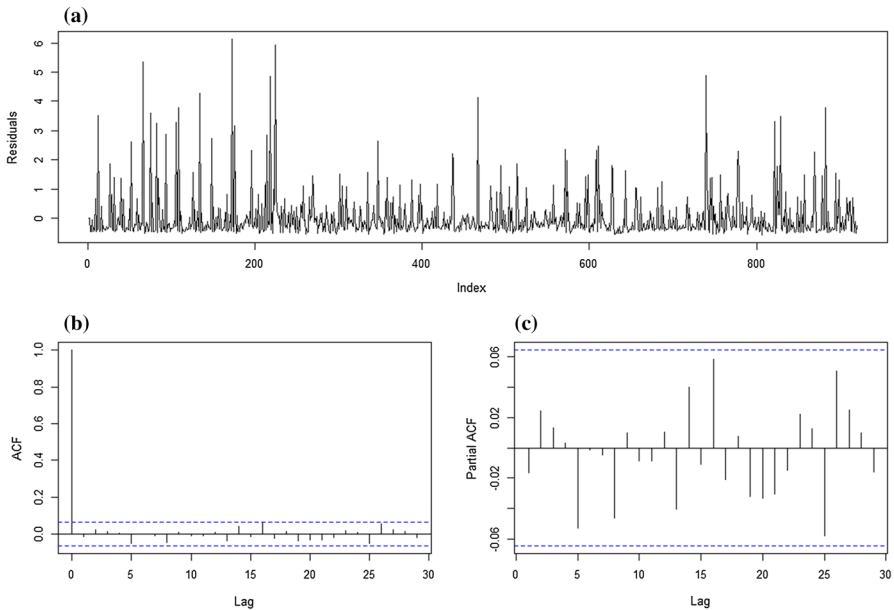
**Table 3** Summary of estimation results for VIX futures tick count data

Par.	BNB-INGARCH(1,1)				NB-INGARCH(1,1)			
	Mean	Std	2.50%	97.50%	Mean	Std	2.50%	97.50%
$\omega$	0.1332	0.0355	0.0835	0.2123	0.2980	0.0715	0.1621	0.4466
$\phi$	0.2145	0.0258	0.1714	0.2659	0.2540	0.0346	0.1896	0.3246
$\tau$	0.7696	0.0308	0.7078	0.8202	0.6793	0.0492	0.5795	0.7700
$r$	3.7093	0.2758	3.1185	4.2318	1.0258	0.0461	0.9424	1.1210
$\alpha$	2.5115	0.1370	2.2315	2.7735				
AIC	7084.96				7328.09			
BIC	7109.09				7347.38			
Par.	BNB-INGARCH(1,2)				NB-INGARCH(1,2)			
	Mean	Std	2.50%	97.50%	Mean	Std	2.50%	97.50%
$\omega$	0.1433	0.0324	0.0880	0.2144	0.2466	0.0538	0.1532	0.3600
$\phi$	0.2372	0.0260	0.1859	0.2869	0.2373	0.0298	0.1799	0.2982
$\tau_1$	0.4302	0.0827	0.2700	0.5964	0.4501	0.0841	0.2840	0.6204
$\tau_2$	0.3160	0.0742	0.1761	0.4620	0.2600	0.0763	0.1059	0.3927
$r$	4.1704	0.5325	3.0340	5.2870	1.0325	0.0477	0.9399	1.1343
$\alpha$	2.5213	0.1557	2.2375	2.8430				
AIC	7069.46				7304.70			
BIC	7098.41				7328.83			
Par.	BNB-INGARCH(2,1)				NB-INGARCH(2,1)			
	Mean	Std	2.50%	97.50%	Mean	Std	2.50%	97.50%
$\omega$	0.1483	0.0395	0.0835	0.2368	0.3574	0.1069	0.1948	0.6661
$\phi_1$	0.1941	0.0298	0.1305	0.2541	0.2180	0.0376	0.1431	0.2898
$\phi_2$	0.0317	0.0268	0.0005	0.1124	0.0910	0.0504	0.0075	0.1973
$\tau$	0.7549	0.0343	0.6846	0.8141	0.6110	0.0818	0.3957	0.7456
$r$	4.2180	0.4933	3.4444	5.1955	1.0279	0.0466	0.9349	1.1182
$\alpha$	2.5060	0.1526	2.2300	2.8121				
AIC	7082.03				7311.75			
BIC	7110.98				7335.87			
Par.	BNB-INGARCH(2,2)				NB-INGARCH(2,2)			
	Mean	Std	2.50%	97.50%	Mean	Std	2.50%	97.50%
$\omega$	0.1670	0.0397	0.1051	0.2561	0.3438	0.0806	0.2054	0.5288
$\phi_1$	0.2123	0.0295	0.1558	0.2786	0.1959	0.0390	0.1239	0.2772
$\phi_2$	0.0737	0.0350	0.0107	0.1474	0.1368	0.0460	0.0473	0.2224
$\tau_1$	0.3096	0.0578	0.2132	0.4382	0.2433	0.1040	0.0604	0.4481
$\tau_2$	0.3866	0.0506	0.2730	0.4679	0.3504	0.0869	0.1810	0.5210
$r$	3.6461	0.2576	3.1299	4.1100	1.0316	0.0480	0.9461	1.1308
$\alpha$	2.5057	0.1497	2.2323	2.8250				
AIC	7073.09				7304.85			
BIC	7106.86				7333.80			





**Fig. 2** Trace plots of the Markov chains for the log-linear BNB-INGARCH(1, 2) model marginal posterior distributions of parameters  $\omega$ ,  $\phi$ ,  $\tau_1$ ,  $\tau_2$ ,  $r$  and  $\alpha$  for VIX futures tick count data. The grey rectangles represent the burn-in periods



**Fig. 3** Diagnostic checking plots for standardized residuals. The top panel reports the standardized Pearson residuals of the log-linear BNB-INGARCH(1, 2) model for VIX futures tick count data. The bottom left and bottom right plots present the auto-correlation and partial auto-correlation functions of the residuals respectively

functions of the standardized residuals for the fitted BNB-INGARCH(1,2) model. The plots indicate no significant serial correlations in the residuals. On the basis of the diagnostic checking figures, the auto-correlation characteristics in the futures tick count data can be captured and described by the dynamics of log-linear BNB-INGARCH models. We conclude that the proposed and fitted model is adequate.

## 6 Conclusion

The Bayesian estimation approach and model selections for the proposed log-linear Beta-negative binomial integer-valued GARCH model have been presented. Parameter estimations for the proposed process are performed based on adaptive Markov chain Monte Carlo methods. The empirical application on high-frequency intraday VIX futures tick count data indicates that the proposed model is adequate and provides better performance than the INGARCH model under negative binomial distribution assumptions. This research, by extending the log-linear BNB-INGARCH model in Bayesian frameworks, complements another aspect of current literature on modelling discrete time series with heavy-tailedness characteristics.

**Acknowledgements** The authors gratefully acknowledge the partial sponsorship offered by OptiRisk Systems for the present investigation and giving us permission to access the VIX futures tick count data.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmad A, Francq C (2016) Poisson QMLE of count time series models. *J Time Ser Anal* 37(3):291–314
- Chan K, Ledolter J (1995) Monte Carlo EM estimation for time series models involving counts. *J Am Stat Assoc* 90:242–252
- Chen CW, Khamthong K (2020) Bayesian modelling of nonlinear negative binomial integer-valued GARCHX models. *Stat Model* 20:537–561
- Chen CW, Lee S (2017) Bayesian causality test for integer-valued time series models with applications to climate and crime data. *J R Stat Soc Ser C* 66:797–814
- Chen CW, So MK (2006) On a threshold heteroscedastic model. *Int J Forecast* 22:73–89
- Chen CW, Khamthong K, Lee S (2019) Markov switching integer-valued generalized auto-regressive conditional heteroscedastic models for dengue counts. *J Roy Stat Soc Ser C (Appl Stat)* 68:963–983
- Chib S, Winkelmann R (2001) Markov chain monte Carlo analysis of correlated count data. *J Bus Econ Stat* 19:428–435

- Christou V, Fokianos K (2014) Quasi-likelihood inference for negative binomial time series models. *J Time Ser Anal* 35:55–78
- Contreras J, Espinola R, Nogales FJ, Conejo AJ (2003) Arima models to predict next-day electricity prices. *IEEE Trans Power Syst* 18:1014–1020
- Cox DR, Gudmundsson G, Lindgren G, Bondesson L, Harsaae E, Laake P, Juselius K, Lauritzen SL (1981) Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scand J Stat* 8:93–115
- Davis RA, Dunsmuir WT, Streett SB (2003) Observation-driven models for Poisson counts. *Biometrika* 90:777–790
- Davis RA, Dunsmuir WT, Streett SB (2005) Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodol Comput Appl Probab* 7:149–159
- Davis RA, Holan SH, Lund R, Ravishanker N (2016) Handbook of discrete-valued time series. CRC Press, Boca Raton
- Douc R, Doukhan P, Moulines E (2013) Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stoch Process Appl* 123(7):2620–2647
- Doukhan P, Fokianos K, Tjøstheim D (2012) On weak dependence conditions for Poisson autoregressions. *Stat Probab Lett* 82:942–948
- Drescher D (2005) Alternative distributions for observation driven count series models. *Economics Working Paper* 2005-11
- Dunsmuir WT, Scott DJ (2015) The GLARMA package for observation-driven time series regression of counts. *J Stat Softw* 67:1–36
- Durbin J, Koopman SJ (2000) Time series analysis of non-gaussian observations based on state space models from both classical and Bayesian perspectives. *J R Stat Soc Ser B (Stat Methodol)* 62:3–56
- Fahrmeir L, Tutz G, Hennevoig W, Salem E (2013) Multivariate statistical modelling based on generalized linear models. Springer Series in Statistics. Springer, New York
- Ferland R, Latour A, Oraichi D (2006) Integer-valued GARCH process. *J Time Ser Anal* 27:923–942
- Fokianos K (2012) 12—count time series models. In: Time series analysis: methods and applications, Handbook of Statistics, vol 30. Elsevier, pp 315–347
- Fokianos K, Tjøstheim D (2011) Log-linear Poisson autoregression. *J Multivar Anal* 102(3):563–578
- Gorgi P (2020) Beta-negative binomial auto-regressions for modelling integer-valued time series with extreme observations. *J R Stat Soc Ser B (Stat Methodol)* 82:1325–1347
- Harvey A, Luati A (2014) Filtering with heavy tails. *J Am Stat Assoc* 109:1112–1122
- Hay JL, Pettitt AN (2001) Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics* 2:433–444
- Jung RC, Kukuk M, Liesenfeld R (2006) Time series of count data: modeling, estimation and diagnostics. *Comput Stat Data Anal* 51:2350–2364
- Kedem B, Fokianos K (2005) Regression models for time series analysis. Wiley Series in Probability and Statistics, Wiley
- Khashei M, Bijari M, Ardali GAR (2009) Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNS). *Neurocomputing* 72:956–967
- Kuk AY, Cheng YW (1997) The Monte Carlo Newton–Raphson algorithm. *J Stat Comput Simul* 59:233–250
- Lee CM, Ko CN (2011) Short-term load forecasting using lifting scheme and ARIMA models. *Expert Syst Appl* 38:5902–5911
- Leroux BG, Puterman ML (1992) Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics* 48:545–558
- Liboschik T, Fokianos K, Fried R (2017) tscount: an R package for analysis of count time series following generalized linear models. *J Stat Softw* 82:1–51
- McCabe BP, Martin GM (2005) Bayesian predictions of low count time series. *Int J Forecast* 21:315–330
- Nariswari R, Pudjihastuti H (2019) Bayesian forecasting for time series of count data. *Procedia Comput Sci* 157:427–435
- Neumann MH (2011) Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* 17:1268–1284
- Qian L, Li Q, Zhu F (2020) Modelling heavy-tailedness in count time series. *Appl Math Model* 82:766–784
- R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

- Scotto MG, Weiß CH, Gouveia S (2015) Thinning-based models in the analysis of integer-valued time series: a review. *Stat Model* 15:590–618
- Silveira de Andrade B, Andrade MG, Ehlers RS (2015) Bayesian GARMA models for count data. *Commun Stat Case Stud Data Anal Appl* 1(4):192–205
- Tjøstheim D (2016) Count time series with observation-driven autoregressive parameter dynamics. In: *Handbook of discrete-valued time series*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods, CRC Press, pp 77–100
- Truong BC, Chen CW, Sriboonchitta S (2017) Hysteretic Poisson INGARCH model for integer-valued time series. *Stat Model* 17:401–422
- Wang C, Liu H, Yao JF, Davis RA, Li WK (2014) Self-excited threshold Poisson autoregression. *J Am Stat Assoc* 109(506):777–787
- Weiß CH (2018) *An introduction to discrete-valued time series*. John Wiley & Sons, Hoboken
- West M, Harrison J (1989) *Bayesian forecasting and dynamic models*
- Winkelmann R (2008) *Econometric analysis of count data*. Springer, Berlin
- Xiong L, Zhu F (2019) Robust quasi-likelihood estimation for the negative binomial integer-valued GARCH (1,1) model with an application to transaction counts. *J Stat Plan Inference* 203:178–198
- Zeger SL (1988) A regression model for time series of counts. *Biometrika* 75:621–629
- Zeger SL, Qaqish B (1988) Markov regression models for time series: a quasi-likelihood approach. *Biometrics* 44:1019–1031
- Zhu F (2011) A negative binomial integer-valued GARCH model. *J Time Ser Anal* 32:54–67

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.