J. Math. Anal. Appl. 389 (2012) 58–71

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models

Fukang Zhu

School of Mathematics, Jilin University, Changchun 130012, China

A R T I C L E I N F O

Article history:
Received 26 June 2011
Available online 19 November 2011
Submitted by U. Stadtmueller

Keywords:
Asymptotics
Ergodicity
Generalized Poisson
Integer-valued GARCH models
Overdispersion
Stationarity
Time series of counts
Underdispersion

A B S T R A C T

Overdispersion in time series of counts is very common and has been well studied by many authors, but the opposite phenomenon of underdispersion may also be encountered in real applications and receives little attention. Based on popularity of the generalized Poisson distribution in regression count models and of Poisson INGARCH models in time series analysis, we introduce a generalized Poisson INGARCH model, which can account for both overdispersion and underdispersion. Compared with the double Poisson INGARCH model, conditions for the existence and ergodicity of such a process are easily given. We analyze the autocorrelation structure and also derive expressions for moments of order 1 and 2. We consider the maximum likelihood estimators for the parameters and establish their consistency and asymptotic normality. We apply the proposed model to one overdispersed real example and one underdispersed real example, respectively, which indicates that the proposed methodology performs better than other conventional model-based methods in the literature.

1. Introduction

Time series of counts are commonly observed in real-world applications, so a number of time series models for counts have been proposed, which are able to describe different types of marginal distribution and autocorrelation structure. See [2,19–21,24,30], among others.

The Poisson distribution provides a standard framework for the analysis of count data, but the requirement that the variance should equal the mean is often too restrictive in practice. Frequently data are overdispersed, with the variance greater than the mean, and there are many alternative distributions that can be used to model the data. The opposite phenomenon of underdispersion, where the variance is less than the mean, occurs less frequently and the choice of distributions is much narrower. However, there are situations in which underdispersion is well documented, see [27] and references therein for some real examples.

As a natural extension of the Poisson distribution, the generalized Poisson (GP) distribution, introduced in [6] as an approximation of a generalized negative binomial distribution and studied extensively by Consul [3] and Consul and Famoye [5], is more flexible and allows for overdispersion or underdispersion. Regression models based on the GP distribution have been studied by many authors, including Consul and Famoye [4], Famoye [9], Wang and Famoye [28], Özmen [26] and Famoye et al. [10].

E-mail address: zfk8010@163.com.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2011.11.042

Recently, Heinen [17] and Ferland et al. [11] proposed an integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) model, which is defined as follows

$$\begin{cases} X_t \mid \mathcal{F}_{t-1} : \mathcal{P}(\lambda_t), \quad \forall t \in \mathbb{Z}, \\ \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \end{cases} \tag{1.1}$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, $p \geqslant 1$, $q \geqslant 0$, and $\mathcal{F}_{t-1}$ is the σ-field generated by $\{X_{t-1}, X_{t-2}, \ldots\}$. Many aspects about this model have been considered, such as, existence [11,16], ergodicity [14,25], estimating methods [14,38,41,42], testing [13,25,40], autocorrelation functions [32], unconditional distributions [33], higher-order moments [34], generalizations [15,17,36,37,39] and real applications [23]. Fokianos [12] reviewed some recent progress on INGARCH models.

Model (1.1) can only deal with overdispersion in time series of counts, but underdispersion may also be encountered in real applications (see Section 5 for real examples). Heinen [17] proposed an INGARCH$(1,1)$ model based on the double Poisson (DP) distribution introduced by Efron [8] to deal with underdispersion, but it is difficult to be utilized because of the intractability of the normalizing constant and moments. In addition, Heinen [17] did not give a real application to underdispersed data. Third, the DP distribution has not been well studied and its many properties remain to be unknown, thus some theoretical aspects of the DP-INGARCH$(1,1)$ model may be difficult to be established. Based on the above considerations, we propose an INGARCH model based on the GP distribution, which can account for both overdispersion and underdispersion. Moreover, we can establish some needed theoretical results easily.

To model underdispersion, the binomial distribution $b(n, p)$ is another alternative, but the parameter n is a discrete-valued parameter and the differentiation with respect to n is problematic, so we can not obtain the joint maximum likelihood estimator (MLE) of n and other parameters. For the similar problem in the negative binomial (NB) INGARCH model, see [36]. In this sense, the GP distribution is a better choice.

The paper is organized as follows. In Section 2, we briefly introduce and review the GP distribution. In Section 3 we describe the GP-INGARCH model, conditions for the existence and ergodicity of such a process are given. We also give a set of equations from which the variance and autocorrelation function can be obtained. We discuss the maximum likelihood estimation procedure and establish asymptotic properties of the estimators in Section 4. In Section 5 we apply the proposed model to one overdispersed real example and one underdispersed real example, respectively, which demonstrates the usefulness and flexibility of the proposed model in fitting time series of counts which do not seem to follow other conventional models in the literature. Section 6 concludes.

2. The generalized Poisson distribution

First, recall the definition of the GP distribution (see, e.g., [5]). A random variable X has a GP distribution with parameters λ and κ, which we denote by $\mathrm{GP}(\lambda, \kappa)$, if its probability mass function is

$$P(X = x) = \begin{cases} \lambda(\lambda + \kappa x)^{x-1} \mathrm{e}^{-(\lambda+\kappa x)}/x!, & x = 0, 1, 2, \ldots, \\ 0, & \text{for } x > m \text{ if } \kappa < 0, \end{cases}$$

where $\lambda > 0$, $\max(-1, -\lambda/m) < \kappa < 1$, and m $(\geqslant 4)$ is the largest positive integer for which $\lambda + \kappa m > 0$ when $\kappa < 0$. When $\kappa < 0$, the distribution includes a truncation due to $P(X = x) = 0$ for all $x > m$ and the sum $\sum_{x=0}^{m} P(X = x)$ is usually a little less than unity. However, this truncation error is less than 0.5% when $m \geqslant 4$ and so the truncation error does not make any difference in practical applications [5, p. 165]. The GP distribution, also known as the Lagrangian Poisson distribution, is a kind of Poisson-stopped-sum distribution. It reduces to the usual Poisson distribution with parameter λ when $\kappa = 0$.

The probability generating function of the GP distribution is

$$g(u) = E\big(u^X\big) = \exp\big\{\lambda(z-1)\big\}, \quad \text{where } z = u \exp\big\{\kappa(z-1)\big\}. \tag{2.1}$$

Alternative representation of the probability generating function is

$$p_X(z) = \exp\big\{-(\kappa/\lambda)\big[W\big(-\lambda z \mathrm{e}^{-\lambda}\big) + \lambda\big]\big\},$$

where W is the Lambert's function defined as $W(x)\exp(W(x)) = x$. By putting $z = \mathrm{e}^s$ and $u = \mathrm{e}^t$ in (2.1), one obtain the moment generating function for the GP distribution as

$$M_X(t) = E\,\mathrm{e}^{tX} = \exp\big\{\lambda\big(\mathrm{e}^s - 1\big)\big\}, \quad \text{where } s = t + \kappa\big(\mathrm{e}^s - 1\big). \tag{2.2}$$

Then the cumulant generating function (cgf) of the GP distribution becomes

$$\kappa_X(t) = \ln M_X(t) = \lambda\big(\mathrm{e}^s - 1\big). \tag{2.3}$$

From (2.2) we know that the sum $X_1 + X_2 + \cdots + X_l$ of l independent GP random variables $X_1, X_2, \ldots, X_l$, with parameters $(\lambda_1, \kappa), (\lambda_2, \kappa), \ldots, (\lambda_l, \kappa)$, respectively, is a GP random variable with parameters $(\lambda_1 + \lambda_2 + \cdots + \lambda_l, \kappa)$, which means that the GP distribution has the additive property like the Poisson distribution. This property will be used several times later.

For the GP distribution $\mathrm{GP}(\lambda,\kappa)$, a recurrence relation between the noncentral moments $\mu_l = E(X^l)$ is

$$(1-\kappa)\mu_{l+1} = \lambda\mu_l + \lambda\frac{\partial\mu_l}{\partial\lambda} + \kappa\frac{\partial\mu_l}{\partial\kappa}, \quad k = 0,1,2,\ldots. \tag{2.4}$$

From $\mu_0 = 1$ and (2.4) we obtain that

$$\mu_1 = \frac{\lambda}{1-\kappa}, \qquad \mu_2 = \frac{\lambda^2}{(1-\kappa)^2} + \frac{\lambda}{(1-\kappa)^3}, \qquad \mu_3 = \frac{\lambda^3}{(1-\kappa)^3} + \frac{3\lambda^2}{(1-\kappa)^4} + \frac{(1+2\kappa)\lambda}{(1-\kappa)^5}, \ldots.$$

All the moments of the GP distribution exist for $\kappa < 1$. First, the variance equals $\lambda/(1-\kappa)^3$, which is greater than, equal to, or less than the mean according to whether $\kappa > 0$, $\kappa = 0$, or $\kappa < 0$, respectively. Second, by induction, we have

$$\mu_l = \sum_{i=0}^{l} a_{li}\lambda^i, \quad l = 1,2,\ldots, \tag{2.5}$$

where a_{li} is not related to λ, and $a_{ll} = 1/(1-\kappa)^l$.

Consul and Shoukri [7] also obtained the negative integer moments of $\mathrm{GP}(\lambda,\kappa)$. In particular, we have

$$E\left(X + \frac{\lambda}{\kappa}\right)^{-1} = \frac{\kappa}{\lambda} - \frac{\kappa^2}{\lambda+\kappa}, \tag{2.6}$$

$$E\left(X + \frac{\lambda}{\kappa}\right)^{-2} = \frac{\kappa^2}{\lambda^2} - \frac{\kappa^3}{\lambda(\lambda+\kappa)} - \frac{\kappa^3}{(\lambda+\kappa)^2} + \frac{\kappa^4}{(\lambda+\kappa)(\lambda+2\kappa)}. \tag{2.7}$$

There is another parametrization of the GP distribution [9,28], which is also known as the Abel distribution. In this paper we just focus on the parametrization (2.1).

3. The generalized Poisson INGARCH model

Let $\{X_t\}$ be a time series of counts. We assume that, conditional on the past information, the random variables $X_1,\ldots,X_n$ are independent, and the conditional distribution of X_t is specified by a GP distribution, i.e.,

$$X_t \mid \mathcal{F}_{t-1} : \mathcal{GP}(\lambda_t^*, \kappa), \qquad \frac{\lambda_t^*}{1-\kappa} = \lambda_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i X_{t-i} + \sum_{j=1}^{q}\beta_j\lambda_{t-j}, \tag{3.1}$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$, $i = 1,\ldots,p$, $j = 1,\ldots,q$, $p \geqslant 1$, $q \geqslant 0$, $\max(-1, -\lambda_t^*/4) < \kappa < 1$, $\mathcal{F}_{t-1}$ is the σ-field generated by $\{X_{t-1}, X_{t-2},\ldots\}$. The above model is denoted by GP-INGARCH(p,q). When $q = 0$, the above model is denoted by GP-INARCH(p). Clearly, when $\kappa = 0$, the model (3.1) reduces to the model (1.1).

In the following theorem we give a sufficient condition under which there exists a stationary GP-INGARCH(p,q) process.

Theorem 1. *If $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1$, then there exists a unique strictly stationary process $\{X_t\}_{t\in\mathbb{Z}}$ that satisfied* (3.1)*. Moreover, the first two moments are finite.*

Proof. The theorem can be proved by using techniques discussed in [11] or [16], here we adopt the former. Let $D(B) = 1 - \beta_1 B - \cdots - \beta_q B^q$, $G(B) = \alpha_1 B + \cdots + \alpha_p B^p$, where B is the backshift operator. Let

$$\lambda_t = D^{-1}(B)\big(\alpha_0 + G(B)X_t\big) = \alpha_0 D^{-1}(1) + H(B)X_t,$$

where $H(B) = D^{-1}(B)G(B) = \sum_{j=1}^{\infty}\psi_j B^j$. Let $\{U_t\}_{t\in\mathbb{Z}}$ be a sequence of independent GP random variables with parameters $(\psi_0 = \alpha_0/D(1), \kappa)$. For each $t \in \mathbb{Z}$ and $i \in \mathbb{Z}^+$, let $\{Z_{t,i,j}\}_{j\in\mathbb{Z}^+}$ represent a sequence of independent GP random variables having parameters (ψ_i, κ). We also assume that all the random variables U_s, $Z_{t,i,j}$ ($s \in \mathbb{Z}$, $t \in \mathbb{Z}$, $i \in \mathbb{Z}^+$ and $j \in \mathbb{Z}^+$) are mutually independent. Define

$$X_t^{(n)} = \begin{cases} 0, & n < 0; \\ (1-\kappa)U_t, & n = 0; \\ (1-\kappa)U_t + (1-\kappa)\sum_{i=1}^{n}\sum_{j=1}^{X_{t-i}^{(n-i)}} Z_{t-i,i,j}, & n > 0. \end{cases} \tag{3.2}$$

Using the thinning operation (see, e.g., [41]), $X_t^{(n)}$ admits the representation

$$X_t^{(n)} = (1-\kappa)U_t + (1-\kappa)\sum_{i=1}^{n}\varphi_i^{(t-i)} \circ X_{t-i}^{(n-i)}, \quad n > 0,$$

where $\varphi_i = \psi_i/(1-\kappa)$. In the above notation $\varphi_i^{(\tau)}\circ$ indicates that the sequence of GP random variables of common mean φ_i involved in the thinning operation are those that correspond to time τ, i.e. the sequence $\{Z_{\tau,i,j}\}_{j\in\mathbb{Z}^+}$.

The expectation and variance of $X_t^{(n)}$ are well defined, because $X_t^{(n)}$ is a finite sum of independent GP random variables. It is easily seen that $E(X_t^{(n)})$ does not depend on t, it just depends on n and will be denoted by μ_n. Using (3.2) and the fact that $\mu_k = 0$ if $k < 0$, we have

$$\mu_n = \psi_0 + \sum_{j=1}^{\infty} \psi_j \mu_{n-j} = D^{-1}(B)\alpha_0 + H(B)\mu_n,$$

then $[D(B) - G(B)]\mu_n = \alpha_0$. Using arguments similar to those in Propositions 2 and 3 of [11] we know that $\{X_t^{(n)}\}_{n\in\mathbb{Z}^+}$ has an almost sure limit X_t and is a strictly stationary process for each n. Then we know that $\{X_t\}_{t\in\mathbb{Z}}$ is a strictly stationary process.

Using arguments similar to those in Proposition 4 of [11] we know that $E(X_t)$ is finite and

$$E\big(X_t^{(n)}\big)^2 \leqslant (1-\kappa)^2\left[E\big(U_t^2\big) + \big[2E(U_t)+1\big]\sum_{i=1}^{n}\varphi_i E(X_t) + \left(\sum_{i=1}^{n}\varphi_i\right)^2 E\big(X_t^{(n)}\big)^2\right].$$

Then

$$\begin{aligned} E\big(X_t^{(n)}\big)^2 &\leqslant \frac{(1-\kappa)^2E(U_t^2) + (1-\kappa)E(X_t)[2E(U_t)+1]\sum_{i=1}^{n}\psi_i}{1-(\sum_{i=1}^{n}\psi_i)^2} \\ &\leqslant \frac{(1-\kappa)^2E(U_t^2) + (1-\kappa)E(X_t)[2E(U_t)+1]\sum_{i=1}^{\infty}\psi_i}{1-(\sum_{i=1}^{\infty}\psi_i)^2} \equiv C. \end{aligned}$$

By the Lebesgue's dominated convergence theorem, we conclude that $E(X_t^2) \leqslant C$. Therefore, the first two moments are finite. $r_t^{(n)}$ therein is replaced by $(1-\kappa)U_t + (1-\kappa)\sum_{i=1}^{n}\sum_{j=1}^{X_{t-i}} Z_{t-i,i,j}$. Using arguments similar to those in Proposition 5 and Section 2.6 of [11] we know that $X_t \mid \mathcal{F}_{t-1} : \mathcal{GP}(\lambda_t^*, \kappa)$. □

The conditional mean and conditional variance of X_t are given by

$$E(X_t \mid \mathcal{F}_{t-1}) = \frac{\lambda_t^*}{1-\kappa} = \lambda_t, \qquad \mathrm{Var}(X_t \mid \mathcal{F}_{t-1}) = \frac{\lambda_t^*}{(1-\kappa)^3} = \phi^2\lambda_t,$$

where $\phi = 1/(1-\kappa)$, then

$$\begin{aligned} &\mu \equiv E(X_t) = \frac{\alpha_0}{1-\sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j}, \\ &\mathrm{Var}(X_t) = E\big(\mathrm{Var}(X_t \mid \mathcal{F}_{t-1})\big) + \mathrm{Var}\big(E(X_t \mid \mathcal{F}_{t-1})\big) = E\big(\phi^2\lambda_t\big) + \mathrm{Var}(\lambda_t) = \phi^2\mu + \mathrm{Var}(\lambda_t). \end{aligned}$$

From Theorem 1 in [32] we know that the following theorem holds, which gives a set of equations from which the variance and autocorrelation function can be obtained.

Theorem 2. *Suppose that $\{X_t\}$ follows the model* (3.1) *with $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1$. Let the autocovariances $\gamma_X(l) = \mathrm{Cov}(X_t, X_{t-l})$, $\gamma_\lambda(l) = \mathrm{Cov}(\lambda_t, \lambda_{t-l})$, then they satisfy the equations*

$$\begin{aligned} \gamma_X(l) &= \sum_{i=1}^{p}\alpha_i\gamma_X\big(|l-i|\big) + \sum_{j=1}^{\min(l-1,q)}\beta_j\gamma_X(l-j) + \sum_{j=l}^{q}\beta_j\gamma_\lambda(j-l), \quad l \geqslant 1; \\ \gamma_\lambda(l) &= \sum_{i=1}^{\min(l,p)}\alpha_i\gamma_\lambda(l-i) + \sum_{i=l+1}^{p}\alpha_i\gamma_X(i-l) + \sum_{j=1}^{q}\beta_j\gamma_\lambda\big(|l-j|\big), \quad l \geqslant 0. \end{aligned}$$

Example 1. Consider the GP-INGARCH(1, 1) model. With arguments similar to those in Example 1 of [32] we have

$$\mathrm{Var}(\lambda_t) = \frac{\phi^2\alpha_1^2\mu}{1-(\alpha_1+\beta_1)^2},$$

then

$$\mathrm{Var}(X_t) = \phi^2\mu + \mathrm{Var}(\lambda_t) = \frac{\phi^2\mu[1-(\alpha_1+\beta_1)^2+\alpha_1^2]}{1-(\alpha_1+\beta_1)^2}.$$

The autocorrelations are given by

$$\rho_\lambda(l) = (\alpha_1+\beta_1)^l, \quad l \geqslant 0;$$
$$\rho_X(l) = (\alpha_1+\beta_1)^{l-1}\frac{\alpha_1[1-\beta_1(\alpha_1+\beta_1)]}{1-(\alpha_1+\beta_1)^2+\alpha_1^2}, \quad l \geqslant 1.$$

Corollary 1. *Suppose that $\{X_t\}$ following the GP-INARCH(p) model is second-order stationary, then the autocovariance function $\gamma_X(l)$ satisfies the equations*

$$\gamma_X(l) = \sum_{i=1}^{p}\alpha_i\gamma_X\big(|l-i|\big), \quad l \geqslant 1.$$

Example 2. Consider the GP-INARCH(1) model. We can obtain the cumulants by using the techniques given in Example 2 of [32], but it is too tedious and complicated for deriving higher-order cumulants. In addition, we can not obtain a recurrent relation for the cumulants because of the complexity of the cgf of the GP distribution. As an illustration, we consider the first two cumulants. In fact, from (2.3) we have

$$\kappa_X(t) = (1-\kappa)\alpha_0\big(\mathrm{e}^s-1\big) + \kappa_X\big((1-\kappa)\alpha_1\big(\mathrm{e}^s-1\big)\big), \quad s = t+\kappa\big(\mathrm{e}^s-1\big),$$

where s is a function of t, and

$$\frac{\partial s}{\partial t} = \frac{1}{1-\kappa\mathrm{e}^s}, \qquad \frac{\partial^2 s}{\partial t^2} = \frac{\kappa\mathrm{e}^s}{(1-\kappa\mathrm{e}^s)^3}.$$

Then

$$\kappa_X'(t) = (1-\kappa)\big[\alpha_0+\alpha_1\kappa_X'\big((1-\kappa)\alpha_1\big(\mathrm{e}^s-1\big)\big)\big]\mathrm{e}^s\frac{\partial s}{\partial t}, \tag{3.3}$$
$$\kappa_X''(t) = \kappa_X'(t)\left[\frac{\partial s}{\partial t}+\frac{\partial^2 s}{\partial t^2}\left(\frac{\partial s}{\partial t}\right)^{-1}\right] + \left[(1-\kappa)\alpha_1\mathrm{e}^s\frac{\partial s}{\partial t}\right]^2\kappa_X''\big((1-\kappa)\alpha_1\big(\mathrm{e}^s-1\big)\big). \tag{3.4}$$

If $t=0$, then $s=0$ because of arbitrariness of κ. Let $\kappa_l = \kappa_X^{(l)}(0)$, then from (3.3) and (3.4) we have

$$\kappa_1 = \alpha_0+\alpha_1\kappa_1,$$
$$\kappa_2 = \kappa_1\left[\frac{1}{1-\kappa}+\frac{\kappa}{(1-\kappa)^2}\right]+\alpha_1^2\kappa_2,$$

thus we obtain the first two cumulants

$$\kappa_1 = \mu = \frac{\alpha_0}{1-\alpha_1}, \qquad \kappa_2 = \mathrm{Var}(X_t) = \frac{\phi^2\alpha_0}{(1-\alpha_1)(1-\alpha_1^2)}.$$

The above results are included in Example 1. In both examples the variance-mean ratio is changed by the factor ϕ^2 compared to the corresponding INGARCH expressions, which implies (i) if $\phi > 1$, then the unconditional overdispersion of the usual INGARCH model is further increased; and (ii) if ϕ is "sufficiently small", then we have unconditional underdispersion.

In what follows we will focus on model (3.1) with $p=q=1$, i.e.,

$$X_t \mid \mathcal{F}_{t-1} : \mathcal{GP}\big(\lambda_t^*, \kappa\big), \qquad \frac{\lambda_t^*}{1-\kappa} = \lambda_t = \alpha_0+\alpha_1X_{t-1}+\beta_1\lambda_{t-1}. \tag{3.5}$$

The following theorem gives geometric ergodicity of the bivariate process $\{(X_t,\lambda_t)\}_{t\in\mathbb{N}}$ defined by (3.5), which is the starting point for establishing the limiting behavior of the MLE discussed in the next section.

Theorem 3. *Suppose that $\{X_t\}$ follows* (3.5) *with $\alpha_1+\beta_1<1$, then $\{(X_t,\lambda_t)\}_{t\in\mathbb{N}}$ is geometrically ergodic.*

The proof of Theorem 3 is done with the same arguments as used by [25] for proving his Theorem 3.1, so we omit the details. Similar to [25], the additive property of the GP distribution is crucial in the proof.

Theorem 4. *Suppose that $\{X_t\}$ follows* (3.5)*, then the moments of X_t are all finite if and only if $\alpha_1+\beta_1<1$.*

Proof. From (2.5) we have

$$E\left(X_t^m \mid \mathcal{F}_{t-1}\right)=\sum_{i=0}^{m} b_{mi}\lambda_t^i,$$

where b_{mi} is not related to λ_t, and $b_{mm}=1$. Notice that

$$\lambda_t^i=(\alpha_0+\alpha_1 X_{t-1}+\beta_1\lambda_{t-1})^i=\sum_{n=0}^{i}\binom{i}{n}\alpha_0^{i-n}\sum_{j=0}^{n}\binom{n}{j}\alpha_1^j\beta_1^{n-j}X_{t-1}^j\lambda_{t-1}^{n-j},$$

thus

$$E\left(\lambda_t^i \mid \mathcal{F}_{t-2}\right)=\sum_{n=0}^{i}\binom{i}{n}\alpha_0^{i-n}\sum_{j=0}^{n}\sum_{k=0}^{j}\binom{n}{j}b_{jk}\alpha_1^j\beta_1^{n-j}\lambda_{t-1}^{n+k-j}.$$

Then the theorem holds by using the same technique given in Proposition 6 of [11]. □

4. Estimation

Let $\theta^*=(\alpha_0,\alpha_1,\beta_1)^\top$, $\theta=(\phi,\theta^{*\top})^\top=(\theta_1,\theta_2,\theta_3,\theta_4)^\top$, where $\phi=1/(1-\kappa)$, and write the true value of θ as $\theta^0=(\phi^0,\alpha_0^0,\alpha_1^0,\beta_1^0)^\top$. Suppose that the observation $X=(X_1,\dots,X_n)$ is generated from the model (3.3). The conditional likelihood function is

$$\prod_{t=2}^{n}\frac{\lambda_t[\lambda_t+(\phi-1)X_t]^{X_t-1}\phi^{-X_t}\exp\{-[\lambda_t+(\phi-1)X_t]/\phi\}}{X_t!},$$

then the log-likelihood is given by

$$l(\theta)=\sum_{t=2}^{n} l_t(\theta)=\sum_{t=2}^{n}\left\{\ln\lambda_t+(X_t-1)\ln\bigl[\lambda_t+(\phi-1)X_t\bigr]-X_t\ln\phi-\frac{\lambda_t+(\phi-1)X_t}{\phi}-\ln(X_t!)\right\}. \tag{4.1}$$

The score function is defined by

$$S_n(\theta)=\frac{\partial l(\theta)}{\partial\theta}=\sum_{t=2}^{n}\frac{\partial l_t(\theta)}{\partial\theta}$$

with

$$\frac{\partial l_t(\theta)}{\partial\phi}=\frac{X_t(X_t-1)}{\lambda_t+(\phi-1)X_t}-\frac{X_t}{\phi}-\frac{X_t-\lambda_t}{\phi^2}, \tag{4.2}$$

$$\frac{\partial l_t(\theta)}{\partial\theta^*}=\left(\frac{X_t-1}{\lambda_t+(\phi-1)X_t}+\frac{1}{\lambda_t}-\frac{1}{\phi}\right)\frac{\partial\lambda_t}{\partial\theta^*},$$

$$\frac{\partial\lambda_t}{\partial\alpha_0}=1+\beta_1\frac{\partial\lambda_{t-1}}{\partial\alpha_0},\qquad \frac{\partial\lambda_t}{\partial\alpha_1}=X_{t-1}+\beta_1\frac{\partial\lambda_{t-1}}{\partial\alpha_1},\qquad \frac{\partial\lambda_t}{\partial\beta_1}=\lambda_{t-1}+\beta_1\frac{\partial\lambda_{t-1}}{\partial\beta_1}. \tag{4.3}$$

The solution of the equation $S_n(\theta)=0$, if it exists, gives the conditional MLE of θ, denoted by $\hat{\theta}$. The Hessian matrix is given by

$$H_n(\theta)=-\sum_{t=2}^{n}\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^\top} \tag{4.4}$$

with

$$\frac{\partial^2 l_t(\theta)}{\partial\phi^2}=-\frac{X_t^2(X_t-1)}{[\lambda_t+(\phi-1)X_t]^2}+\frac{X_t}{\phi^2}+\frac{2(X_t-\lambda_t)}{\phi^3}, \tag{4.5}$$

$$\frac{\partial^2 l_t(\theta)}{\partial\phi\partial\theta^*}=-\left(\frac{X_t(X_t-1)}{[\lambda_t+(\phi-1)X_t]^2}-\frac{1}{\phi^2}\right)\frac{\partial\lambda_t}{\partial\theta^*}, \tag{4.6}$$

$$\frac{\partial^2 l_t(\theta)}{\partial\theta^*\partial\theta^{*\top}}=-\left(\frac{X_t-1}{[\lambda_t+(\phi-1)X_t]^2}+\frac{1}{\lambda_t^2}\right)\frac{\partial\lambda_t}{\partial\theta^*}\frac{\partial\lambda_t}{\partial\theta^{*\top}}+\left(\frac{X_t-1}{\lambda_t+(\phi-1)X_t}+\frac{1}{\lambda_t}-\frac{1}{\phi}\right)\frac{\partial^2\lambda_t}{\partial\theta^*\partial\theta^{*\top}}, \tag{4.7}$$

$$\frac{\partial^2\lambda_t}{\partial\alpha_0^2}=0,\qquad \frac{\partial^2\lambda_t}{\partial\alpha_1^2}=0,\qquad \frac{\partial^2\lambda_t}{\partial\alpha_0\partial\alpha_1}=0,\qquad \frac{\partial^2\lambda_t}{\partial\alpha_0\partial\beta_1}=\frac{\partial\lambda_{t-1}}{\partial\alpha_0}+\beta_1\frac{\partial^2\lambda_{t-1}}{\partial\alpha_0\partial\beta_1},$$
$$\frac{\partial^2\lambda_t}{\partial\alpha_1\partial\beta_1}=\frac{\partial\lambda_{t-1}}{\partial\alpha_1}+\beta_1\frac{\partial^2\lambda_{t-1}}{\partial\alpha_1\partial\beta_1},\qquad \frac{\partial^2\lambda_t}{\partial\beta_1^2}=2\frac{\partial\lambda_{t-1}}{\partial\beta_1}+\beta_1\frac{\partial^2\lambda_{t-1}}{\partial\beta_1^2}.$$

Due to space limitations, the conditional expectations of terms in (4.2), (4.3) and (4.5)–(A.2) are deferred to Appendix A.

Now we want to establish the asymptotic properties of the MLE $\hat{\theta}$. For this purpose, we make the following two assumptions.

Assumption 1. The parameter space is

$$O\left(\theta^0\right)=\left\{\theta \;\middle|\; 0<\phi_L\leqslant\phi\leqslant\phi_U,\ 0<\delta_L\leqslant\alpha_0\leqslant\delta_U,\ 0<\alpha_L\leqslant\alpha_1\leqslant\alpha_U,\ 0<\beta_L\leqslant\beta_1\leqslant\beta_U,\ \delta_L+2(\phi_L-1)>0\right\}.$$

Assumption 2. $\lambda_t+(\phi-1)X_t\geqslant\omega>0$ for all X_t.

Remark 1. By Assumption 1 we know that $\lambda_t+2(\phi-1)>0$ $(t=1,\dots,n)$, then some necessary results (for example, see the conditional expectation in (A.17)) can remain valid when $\phi<1$. From the log-likelihood given in (4.1) we know that $\lambda_t+(\phi-1)X_t>0$ for all X_t, so it is not awkward to make Assumption 2.

The lower bounds in Assumptions 1 and 2 are only for the technical reason in the proof of Theorem 1. In practice, we can select them to be very close to 0.

Theorem 5. *Consider model* (3.3) *with the true value θ^0 and suppose that $\alpha_1^0+\beta_1^0<1$ and Assumptions* 1 *and* 2 *hold, then there exists a fixed open neighborhood $O(\theta^0)$ of θ^0 such that with probability tending to* 1*, as $n\to\infty$, the log-likelihood function* (4.1) *has a unique maximum point $\hat{\theta}$. Furthermore, $\hat{\theta}$ is consistent and asymptotically normal,*

$$\sqrt{n}\left(\hat{\theta}-\theta^0\right)\xrightarrow{d}\mathcal{N}\left(0,G^{-1}\right),$$

where G is defined in the proof.

Proof. The theorem can be proved with arguments similar to those for proving Theorem 1 in [22] and Theorem 3 in [41], which utilizes Theorems 4.1.1 and 4.1.3 in [1]. Here we introduce the techniques in [14], which takes advantage of the fact that the log-likelihood function is three times differentiable.

First, using the results in (A.3) and (A.4), we know that $\partial l_t(\theta)/\partial\theta$ given in (4.2) and (4.3) is a martingale difference sequence with respect to $\mathcal{F}_{t-1}$. In the following, the conditions of Lemma 1 in [18] will be verified. Using the results in (A.8) and (A.17), we have

$$E\left(\left[\frac{X_t-1}{\lambda_t+(\phi-1)X_t}+\frac{1}{\lambda_t}-\frac{1}{\phi}\right]^2\;\middle|\;\mathcal{F}_{t-1}\right)=\frac{1}{\phi\lambda_t}-\frac{\phi-1}{\phi^2[\lambda_t+2(\phi-1)]}\leqslant\frac{1}{\phi\lambda_t}+\frac{\phi+1}{\phi^2[\lambda_t+2(\phi-1)]}$$
$$\leqslant\frac{1}{\phi_L\delta_L}+\frac{\phi_U+1}{\phi_L^2[\delta_L+2(\phi_L-1)]},$$
$$E\left(\left[\frac{X_t(X_t-1)}{\lambda_t+(\phi-1)X_t}-\frac{X_t}{\phi}-\frac{X_t-\lambda_t}{\phi^2}\right]^2\;\middle|\;\mathcal{F}_{t-1}\right)=\frac{2}{\phi^2}-\frac{4(\phi-1)}{\phi^2[\lambda_t+2(\phi-1)]}\leqslant\frac{2}{\phi^2}+\frac{4(\phi+1)}{\phi^2[\lambda_t+2(\phi-1)]}$$
$$\leqslant\frac{2}{\phi_L^2}+\frac{4(\phi_U+1)}{\phi_L^2[\delta_L+2(\phi_L-1)]}.$$

Fokianos et al. [14] showed that $E\|\partial\lambda_t/\partial\theta^*\|<\infty$, thus $E\|\partial l_t(\theta)/\partial\theta\|<\infty$. Using the same arguments given in Lemma 3.2 of [14], we know that

$$\frac{1}{\sqrt{n}}S_n(\theta)\xrightarrow{d}\mathcal{N}(0,G),$$

where

$$G=E\left(G_t^*(\theta)\right),\qquad G_t^*(\theta)=\operatorname{Var}\left(\frac{\partial l_t(\theta)}{\partial\theta}\;\middle|\;\mathcal{F}_{t-1}\right)=\begin{pmatrix}G_{11} & G_{21}^\top\\ G_{21} & G_{22}\end{pmatrix}$$

with

$$G_{11} = \frac{2\lambda_t}{\phi^2[\lambda_t + 2(\phi - 1)]},$$
$$G_{21} = \frac{-2(\phi - 1)}{\phi^2[\lambda_t + 2(\phi - 1)]} \frac{\partial \lambda_t}{\partial \theta^*},$$
$$G_{22} = \left[\frac{1}{\phi \lambda_t} - \frac{\phi - 1}{\phi^2[\lambda_t + 2(\phi - 1)]}\right] \frac{\partial \lambda_t}{\partial \theta^*} \frac{\partial \lambda_t}{\partial \theta^{*\top}}.$$

Using the results in (A.8), (A.12) and (A.17), we know that

$$E\left(\frac{\partial l_t(\theta)}{\partial \theta} \frac{\partial l_t(\theta)}{\partial \theta^\top}\right) = E\left(-\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^\top}\right).$$

Following the arguments given in Lemma 3.3 of [14], we have

$$\frac{1}{n} H_n(\theta) \xrightarrow{P} G.$$

Notice that the third derivatives are given by

$$\frac{\partial^3 l_t(\theta)}{\partial \phi^3} = \frac{2X_t^3(X_t - 1)}{[\lambda_t + (\phi - 1)X_t]^3} - \frac{2X_t}{\phi^3} - \frac{6(X_t - \lambda_t)}{\phi^4},$$
$$\frac{\partial^3 l_t(\theta)}{\partial \phi^2 \partial \theta^*} = \left(\frac{X_t^2(X_t - 1)}{[\lambda_t + (\phi - 1)X_t]^3} - \frac{2}{\phi^3}\right) \frac{\partial \lambda_t}{\partial \theta^*},$$
$$\frac{\partial^3 l_t(\theta)}{\partial \phi \partial \theta^* \partial \theta^{*\top}} = \frac{2X_t(X_t - 1)}{[\lambda_t + (\phi - 1)X_t]^3} \frac{\partial \lambda_t}{\partial \theta^*} \frac{\partial \lambda_t}{\partial \theta^{*\top}} - \left(\frac{X_t(X_t - 1)}{[\lambda_t + (\phi - 1)X_t]^2} - \frac{1}{\phi^2}\right) \frac{\partial^2 \lambda_t}{\partial \theta^* \partial \theta^{*\top}},$$
$$\begin{aligned}\frac{\partial^3 l_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} = &-\left(\frac{X_t - 1}{[\lambda_t + (\phi - 1)X_t]^2} + \frac{1}{\lambda_t^2}\right)\left(\frac{\partial^2 \lambda_t}{\partial \theta_i \partial \theta_j} \frac{\partial \lambda_t}{\partial \theta_k} + \frac{\partial^2 \lambda_t}{\partial \theta_i \partial \theta_k} \frac{\partial \lambda_t}{\partial \theta_j} + \frac{\partial^2 \lambda_t}{\partial \theta_j \partial \theta_k} \frac{\partial \lambda_t}{\partial \theta_i}\right) \\ &+ 2\left(\frac{X_t - 1}{[\lambda_t + (\phi - 1)X_t]^3} + \frac{1}{\lambda_t^3}\right) \frac{\partial \lambda_t}{\partial \theta_i} \frac{\partial \lambda_t}{\partial \theta_j} \frac{\partial \lambda_t}{\partial \theta_k} + \left(\frac{X_t - 1}{\lambda_t + (\phi - 1)X_t} + \frac{1}{\lambda_t} - \frac{1}{\phi}\right) \frac{\partial^3 \lambda_t}{\partial \theta_i \partial \theta_j \partial \theta_k},\end{aligned}$$
$$i, j, k = 2, 3, 4.$$

It is simple to see that all terms that do not contain the partial derivatives of λ_t can be controlled, such as,

$$\left|\frac{2X_t^3(X_t - 1)}{[\lambda_t + (\phi - 1)X_t]^3} - \frac{2X_t}{\phi^3} - \frac{6(X_t - \lambda_t)}{\phi^4}\right| \leqslant \frac{2X_t^3(X_t + 1)}{\omega^3} + \frac{2X_t}{\phi_L^3} + \frac{6(X_t + \lambda_t)}{\phi_L^4},$$
$$\left|\frac{X_t - 1}{[\lambda_t + (\phi - 1)X_t]^2} + \frac{1}{\lambda_t^2}\right| \leqslant \frac{X_t + 1}{\omega^2} + \frac{1}{\delta_L^2}.$$

Following the arguments given in Lemma 3.4 of [14], we obtain that

$$\max_{i,j,k=1,2,3,4} \sup_{\theta \in O(\theta^0)} \left|\frac{1}{n} \sum_{t=2}^{n} \frac{\partial^3 l_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}\right| \leqslant M_n,$$

and $M_n \xrightarrow{P} M$, where M_n is defined analogously to that in Lemma 3.4 of [14] and M is a finite constant.

Now all the conditions of Lemma 1 in [18] have been verified, thus the theorem holds. $\square$

From [35] we know that the standard errors of MLE $\hat{\theta}$ can be computed from the robust sandwich matrix $H_n^{-1}(\hat{\theta}) S_n(\hat{\theta}) H_n^{-1}(\hat{\theta})$, where $H_n(\theta)$ is given in (4.4) and

$$S_n(\theta) = \sum_{t=2}^{n} \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta^\top}.$$

5. Real data examples

In this section, we discuss possible applications of the introduced GP-INGARCH model. Searching for the maximizer of the log-likelihood function is implemented in `Matlab` by using the constrained nonlinear optimization function `fmincon`. Here the constrained conditions are $\alpha_0 > 0$ and the first-order stationary condition.

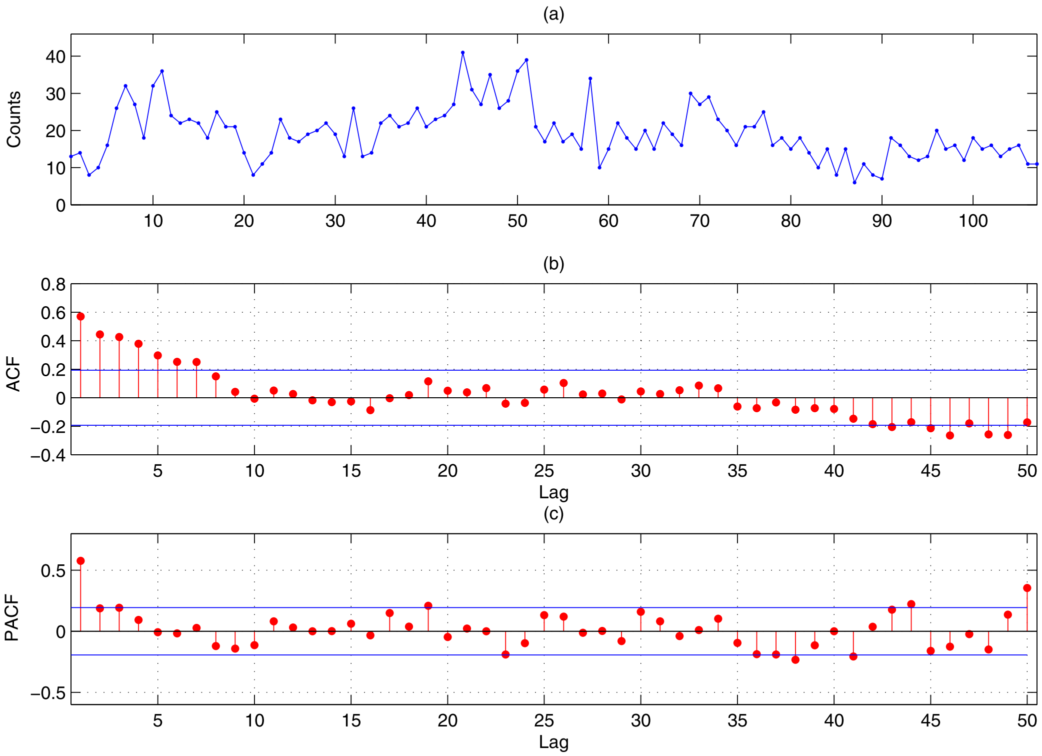

Fig. 1. Major earthquakes counts series: the time plot, the sample autocorrelation and partial autocorrelation function.

5.1. Modeling overdispersion

We use one real example to show good performance of the GP model in modeling overdispersed count data, which is the series of annual counts of major earthquakes (magnitude 7 and above) for the years 1900–2006. These data were originally analyzed by Zucchini and MacDonald [43]. The original series, the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the series are plotted in Fig. 1. The series displays strong positive serial dependence. Empirical mean and variance of the data are given by 19.3645 and 51.5734, respectively, indicating that the true marginal distribution is considerably overdispersed. The sample first-order autocorrelation coefficient (FOAC) is 0.5699. The PACF suggests that INARCH(1) or INGARCH(1, 1) models may be a good choice.

Although Zucchini and MacDonald [43] analyzed these data by virtue of hidden Markov models, we tried to fit them by using the GP-INARCH(1) and GP-INGARCH(1, 1) models. For comparison, we also consider the Poisson models [11], the DP models [17] and the NB models [36]. For more details about estimating the latter two models, one can see [36]. We fitted the NB models with $r = 1, 2, \dots, 10$, we found that $\hat{r} = 10$. We extended the range of r to 15, then we found that $\hat{r} = 15$, which shows that the NB model is not appropriate for these data in some sense. In addition, the interpretation of the estimated parameter $\hat{r} = 15$ is difficult. Parameter estimates and their asymptotic standard errors for other models are summarized in Table 1. The values of Akaike information criterion (AIC) and Bayesian information criterion (BIC) are also provided.

Based on AIC and BIC, we find that INGARCH(1, 1) models give better fit than INARCH(1) models. Within the three fitted INGARCH(1, 1) models, the mean, variance and FOAC are summarized in Table 2. Three models has good performances in fitting mean and FOAC, while only GP and DP model give reasonable variance. Based on the above consideration, we think that the GP-INGARCH(1, 1) model is the best choice for these data.

The set of hypotheses, $H_0 : \phi = 1$ vs. $H_1 : \phi \neq 1$, ask whether the use of Poisson INGARCH is reasonable versus the alternative of fitting GP-INGARCH. We use the following likelihood ratio test statistic:

$$\mathrm{LRT} = -2 \ln \frac{L(\hat{\theta}^{*})}{L(\hat{\theta}^{\star})},$$

where $L(\hat{\theta}^{*})$ is the likelihood function for the Poisson model and $L(\hat{\theta}^{\star})$ is the likelihood function for the GP model. The unknown parameters in each case are estimated by the method of maximum likelihood. Under the null hypothesis, LRT is approximately chi-square distributed with 1 degree of freedom. We compute LTR to be 14.9692. On comparing $\chi^2_{0.99}(1) = 6.6349$, we notice that LRT is significant. Thus, we can conclude that the GP model is more appropriate for this time series.

Table 1
Major earthquakes counts series: parameter estimates with Poisson, DP- and GP-INARCH(1) and INGARCH(1, 1) models. Standard errors are shown in parentheses.

	Model	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}/\hat{\phi}$	AIC	BIC
INARCH(1)	Poisson	7.9476	0.5903			688.5488	693.8567
		(2.5023)	(0.0072)				
	DP	7.9476	0.5903		0.5836	672.0032	679.9364
		(2.5023)	(0.0072)		(0.0079)		
	GP	8.0600	0.5845		1.3088	672.0228	679.9560
		(2.6063)	(0.0076)		(0.0100)		
INGARCH(1, 1)	Poisson	2.6516	0.4057	0.4572		679.7366	687.6697
		(2.6118)	(0.0104)	(0.0224)			
	DP	2.6516	0.4057	0.4572	0.6205	667.4976	678.0365
		(2.6118)	(0.0104)	(0.0224)	(0.0077)		
	GP	2.5837	0.4008	0.4656	1.2739	666.7674	677.3063
		(2.2700)	(0.0092)	(0.0186)	(0.0080)		

Table 2
Major earthquakes counts series: mean, variance and FOAC under the fitted Poisson, DP and GP-INGARCH(1, 1) models.

Model	Poisson	DP	GP
Mean	19.3406	19.3406	19.3391
Variance	31.8045	51.2563	51.6024
FOAC	0.5849	0.5849	0.5832

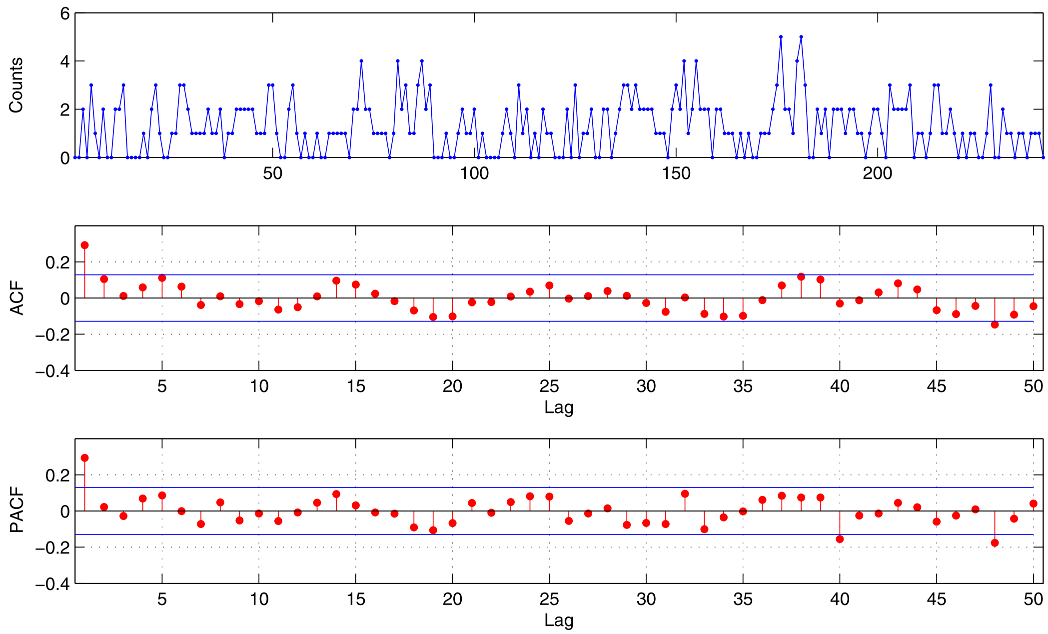

Fig. 2. IP counts series: the time plot, the sample autocorrelation and partial autocorrelation function.

5.2. Modeling underdispersion

To display the flexibility and elegance of the proposed model in modeling underdispersion, we consider one real example, which is the number of different IP addresses (≈ different users) registered within periods of 2-min length at the server of the Department of Statistics of the University of Würzburg in November and December 2005. In particular, we focus on the time series collected on November 29th, 2005, between 10 o'clock in the morning and 6 o'clock in the evening, a time series of length 241. These data have been investigated by Weiß [29,31]. To give an idea about the data structure, Fig. 2 shows the original series, the ACF and PACF of the series. The sample mean and variance are 1.2863 and 1.2052, respectively, which indicates that the data are underdispersed. The sample FOAC is 0.2925.

In [29] the Poisson INAR(1) model was originally proposed for these data, thus it is the natural benchmark model. So we apply the Poisson INAR(1) model, the Poisson, DP- and GP-INARCH(1) and INGARCH(1, 1) models to the data. Parameter estimates and their asymptotic standard errors are summarized in Table 3. The AIC and BIC values are also provided.

Table 3
IP counts series: parameter estimates with Poisson INAR(1), Poisson, DP- and GP-INARCH(1) and INGARCH(1, 1) models. Standard errors are shown in parentheses.

	Model	Par. 1	Par. 2	Par. 3	Par. 4	AIC	BIC
INAR(1)	Poisson	0.2921	0.9143			673.6132	680.5828
	(α, λ)	(0.0614)	(0.0948)				
INARCH(1)	Poisson	0.9265	0.2827			675.1979	682.1508
	(α_0, α_1)	(0.0083)	(0.0038)				
	DP	0.9265	0.2827	0.9623		677.0184	687.4352
	$(\alpha_0, \alpha_1, \gamma)$	(0.0083)	(0.0038)	(0.0046)			
	GP	0.9255	0.2834	0.9157		673.7768	684.1936
	$(\alpha_0, \alpha_1, \phi)$	(0.0081)	(0.0037)	(0.0014)			
INGARCH(1, 1)	Poisson	0.8810	0.2792	0.0388		677.1730	687.5898
	$(\alpha_0, \alpha_1, \beta_1)$	(0.0735)	(0.0045)	(0.0521)			
	DP	0.8810	0.2792	0.0388	0.9624	678.9944	692.8667
	$(\alpha_0, \alpha_1, \beta_1, \gamma)$	(0.0735)	(0.0045)	(0.0521)	(0.0046)		
	GP	0.8806	0.2802	0.0382	0.9156	675.7482	689.6205
	$(\alpha_0, \alpha_1, \beta_1, \phi)$	(0.0709)	(0.0043)	(0.0500)	(0.0014)		

Table 4
IP counts series: mean, variance and FOAC under the fitted Poisson INAR(1), Poisson and GP-INARCH(1) models.

Model	Poisson INAR(1)	Poisson INARCH(1)	GP-INARCH(1)
Mean	1.2916	1.2916	1.2915
Variance	1.2916	1.4038	1.1775
FOAC	0.2921	0.2827	0.2834

Note that we have assumed that $\lambda_t + (\phi - 1)X_t > 0$ when deriving the MLEs for GP models, in other words, the range of X_t is limited to $\{0, 1, \dots, m\}$ with m being large enough. So we need to check the validity of this assumption because $\hat{\phi}$ is less than 1 for this example. For the GP-INARCH(1) model, the minimum of $\{\lambda_t + (\phi - 1)X_t,\ t = 1, 2, \dots, 241\}$ is 0.6727; while for the GP-INGARCH(1, 1) model, the minimum is 0.6740. Thus estimators for GP models are valid. According to comments in Section 2, for practical applications, this restriction is satisfied in most cases.

Based on AIC, we find that the Poisson INAR(1) and GP-INARCH(1) models are the best ones. Based on BIC, we find that the Poisson INAR(1), Poisson and GP-INARCH(1) models are the best ones. Within these three fitted models, the mean, variance and FOAC are summarized in Table 4. All three models exhibit good fits of mean and FOAC, but only the GP-INARCH(1) model gives a reasonable fit of variance, which matches the underdispersed feature. Based on this fact, we conclude that the GP model captures more characteristics of these data.

6. Conclusion

In this paper we introduce a GP-INGARCH model to account for both overdispersion and underdispersion. Conditions for the existence and ergodicity of a GP-INGARCH process are given. The autocorrelation structure is analyzed and expressions for moments of order 1 and 2 are also derived. The maximum likelihood estimators for the parameters are considered and asymptotic properties of the estimators are established. We apply the proposed model to two real examples, which shows that the proposed model not only has good performance in modeling overdispersed data but also has ability to model the underdispersed phenomenon.

Acknowledgments

I gratefully acknowledge comments from the referee that led to a much improved version of this paper. I wish to thank Prof. Christian H. Weiß, Darmstadt University of Technology, for contributing the IP counts data.

This work is supported by National Natural Science Foundation of China (Nos. 11001105, 10971081), Specialized Research Fund for the Doctoral Program of Higher Education (No. 20090061120037), Fundamental Research Fund of Jilin University (Nos. 200903278, 201100011, 200810024), and 985 Project of Jilin University.

Appendix A. Some conditional expectations

Part 1.
By (2.6) and (2.7) we have

$$E\left(\frac{1}{\lambda_t + (\phi - 1)X_t} \,\middle|\, \mathcal{F}_{t-1}\right) = \frac{1}{\lambda_t} - \frac{\phi - 1}{\phi[\lambda_t + (\phi - 1)]}, \tag{A.1}$$

$$E\left(\frac{1}{[\lambda_t+(\phi-1)X_t]^2}\,\middle|\,\mathcal{F}_{t-1}\right)=\frac{1}{\lambda_t^2}-\frac{\phi-1}{\phi\lambda_t[\lambda_t+(\phi-1)]}-\frac{\phi-1}{\phi[\lambda_t+(\phi-1)]^2}+\frac{(\phi-1)^2}{\phi^2[\lambda_t+(\phi-1)][\lambda_t+2(\phi-1)]}. \tag{A.2}$$

Part 2.
By (A.1) we have

$$E\left(\frac{X_t-1}{\lambda_t+(\phi-1)X_t}\,\middle|\,\mathcal{F}_{t-1}\right)=E\left(\frac{1}{\phi-1}\left[1-\frac{\lambda_t+(\phi-1)}{\lambda_t+(\phi-1)X_t}\right]\,\middle|\,\mathcal{F}_{t-1}\right)=\frac{1}{\phi}-\frac{1}{\lambda_t}, \tag{A.3}$$

then from (A.3) we have

$$E\left(\frac{X_t(X_t-1)}{\lambda_t+(\phi-1)X_t}\,\middle|\,\mathcal{F}_{t-1}\right)=E\left(\frac{1}{\phi-1}\left[X_t-1-\frac{\lambda_t(X_t-1)}{\lambda_t+(\phi-1)X_t}\right]\,\middle|\,\mathcal{F}_{t-1}\right)=\frac{\lambda_t}{\phi}. \tag{A.4}$$

Part 3.
By (A.1) and (A.2) we have

$$\begin{aligned}E\left(\frac{X_t^2}{[\lambda_t+(\phi-1)X_t]^2}\,\middle|\,\mathcal{F}_{t-1}\right)&=E\left(\frac{1}{(\phi-1)^2}\left[1-\frac{\lambda_t}{\lambda_t+(\phi-1)X_t}\right]^2\,\middle|\,\mathcal{F}_{t-1}\right)\\&=E\left(\frac{1}{(\phi-1)^2}\left[1-\frac{2\lambda_t}{\lambda_t+(\phi-1)X_t}+\frac{\lambda_t^2}{[\lambda_t+(\phi-1)X_t]^2}\right]\,\middle|\,\mathcal{F}_{t-1}\right)\\&=\frac{\lambda_t}{\phi[\lambda_t+(\phi-1)]^2}+\frac{\lambda_t^2}{\phi^2[\lambda_t+(\phi-1)][\lambda_t+2(\phi-1)]},\end{aligned} \tag{A.5}$$

$$\begin{aligned}E\left(\frac{(X_t-1)^2}{[\lambda_t+(\phi-1)X_t]^2}\,\middle|\,\mathcal{F}_{t-1}\right)&=E\left(\frac{1}{(\phi-1)^2}\left[1-\frac{\lambda_t+(\phi-1)}{\lambda_t+(\phi-1)X_t}\right]^2\,\middle|\,\mathcal{F}_{t-1}\right)\\&=E\left(\frac{1}{(\phi-1)^2}\left[1-\frac{2[\lambda_t+(\phi-1)]}{\lambda_t+(\phi-1)X_t}+\frac{[\lambda_t+(\phi-1)]^2}{[\lambda_t+(\phi-1)X_t]^2}\right]\,\middle|\,\mathcal{F}_{t-1}\right)\\&=\frac{1}{\lambda_t^2}-\frac{1}{\phi\lambda_t}+\frac{1}{\phi^2}-\frac{\phi-1}{\phi^2[\lambda_t+2(\phi-1)]}.\end{aligned} \tag{A.6}$$

Using the fact that $2(X_t-1)=X_t^2-(X_t-1)^2-1$, then from (A.2), (A.5) and (A.6) we have

$$E\left(\frac{X_t-1}{[\lambda_t+(\phi-1)X_t]^2}\,\middle|\,\mathcal{F}_{t-1}\right)=\frac{1}{\phi\lambda_t}-\frac{1}{\lambda_t^2}-\frac{\phi-1}{\phi^2[\lambda_t+2(\phi-1)]}. \tag{A.7}$$

By (A.3), (A.6) and (A.7), it is easy to verify that

$$\begin{aligned}E\left(\left[\frac{X_t-1}{\lambda_t+(\phi-1)X_t}+\frac{1}{\lambda_t}-\frac{1}{\phi}\right]^2\,\middle|\,\mathcal{F}_{t-1}\right)&=E\left(\left(\frac{X_t-1}{[\lambda_t+(\phi-1)X_t]^2}+\frac{1}{\lambda_t^2}\right)\,\middle|\,\mathcal{F}_{t-1}\right)\\&=\frac{1}{\phi\lambda_t}-\frac{\phi-1}{\phi^2[\lambda_t+2(\phi-1)]}.\end{aligned} \tag{A.8}$$

Part 4.
By (A.3) and (A.7) we have

$$\begin{aligned}E\left(\frac{(X_t-1)^3}{[\lambda_t+(\phi-1)X_t]^2}\,\middle|\,\mathcal{F}_{t-1}\right)&=E\left(\frac{(X_t-1)}{(\phi-1)^2}\left[1-\frac{\lambda_t+(\phi-1)}{\lambda_t+(\phi-1)X_t}\right]^2\,\middle|\,\mathcal{F}_{t-1}\right)\\&=E\left(\frac{1}{(\phi-1)^2}\left[1-2\big[\lambda_t+(\phi-1)\big]\frac{(X_t-1)}{\lambda_t+(\phi-1)X_t}\right.\right.\\&\quad\left.\left.+\big[\lambda_t+(\phi-1)\big]^2\frac{(X_t-1)}{[\lambda_t+(\phi-1)X_t]^2}\right]\,\middle|\,\mathcal{F}_{t-1}\right)\\&=\frac{\lambda_t}{\phi^2}+\frac{1}{\phi\lambda_t}-\frac{1}{\lambda_t^2}-\frac{\phi-1}{\phi^2[\lambda_t+2(\phi-1)]}.\end{aligned} \tag{A.9}$$

Using the fact that $X_t(X_t-1)^2=(X_t-1)^2+(X_t-1)^3$, then from (A.6) and (A.9) we have

$$E\left(\frac{X_t(X_t-1)^2}{[\lambda_t+(\phi-1)X_t]^2}\,\middle|\,\mathcal{F}_{t-1}\right)=\frac{\lambda_t+1}{\phi^2}-\frac{2(\phi-1)}{\phi^2[\lambda_t+2(\phi-1)]}. \tag{A.10}$$

Using the fact that $X_t(X_t-1)=(X_t-1)+(X_t-1)^2$, then from (A.6) and (A.7) we have

$$E\left(\frac{X_t(X_t-1)}{[\lambda_t+(\phi-1)X_t]^2}\,\middle|\,\mathcal{F}_{t-1}\right)=\frac{1}{\phi^2}-\frac{2(\phi-1)}{\phi^2[\lambda_t+2(\phi-1)]}. \tag{A.11}$$

By (A.4), (A.10) and (A.11), it is easy to verify that

$$\begin{aligned}&E\left(\left[\frac{X_t(X_t-1)}{\lambda_t+(\phi-1)X_t}-\frac{X_t}{\phi}-\frac{X_t-\lambda_t}{\phi^2}\right]\left[\frac{X_t-1}{\lambda_t+(\phi-1)X_t}+\frac{1}{\lambda_t}-\frac{1}{\phi}\right]\,\middle|\,\mathcal{F}_{t-1}\right)\\&\quad=E\left(\left(\frac{X_t(X_t-1)}{[\lambda_t+(\phi-1)X_t]^2}-\frac{1}{\phi^2}\right)\,\middle|\,\mathcal{F}_{t-1}\right)=-\frac{2(\phi-1)}{\phi^2[\lambda_t+2(\phi-1)]}.\end{aligned} \tag{A.12}$$

Part 5.

Using the fact that $(X_t-1)^2=X_t(X_t-1)-(X_t-1)$, then from (A.3) and (A.4) we have

$$E\left(\frac{(X_t-1)^2}{\lambda_t+(\phi-1)X_t}\,\middle|\,\mathcal{F}_{t-1}\right)=\frac{\lambda_t-1}{\phi}+\frac{1}{\lambda_t}, \tag{A.13}$$

then from (A.6) and (A.13) we have

$$\begin{aligned}&E\left(\frac{X_t^2(X_t-1)^2}{[\lambda_t+(\phi-1)X_t]^2}\,\middle|\,\mathcal{F}_{t-1}\right)\\&\quad=E\left(\frac{(X_t-1)^2}{(\phi-1)^2}\left[1-\frac{\lambda_t}{\lambda_t+(\phi-1)X_t}\right]^2\,\middle|\,\mathcal{F}_{t-1}\right)\\&\quad=E\left(\frac{1}{(\phi-1)^2}\left[(X_t-1)^2-\frac{2\lambda_t(X_t-1)^2}{\lambda_t+(\phi-1)X_t}+\frac{\lambda_t^2(X_t-1)^2}{[\lambda_t+(\phi-1)X_t]^2}\right]\,\middle|\,\mathcal{F}_{t-1}\right)\\&\quad=\frac{\lambda_t^2}{\phi^2}+\frac{(\phi+1)^2\lambda_t}{\phi^2}+\frac{2\lambda_t}{\phi^2[\lambda_t+2(\phi-1)]}.\end{aligned} \tag{A.14}$$

By (A.4) we have

$$E\left(\frac{X_t^2(X_t-1)}{\lambda_t+(\phi-1)X_t}\,\middle|\,\mathcal{F}_{t-1}\right)=E\left(\frac{1}{\phi-1}\left[X_t^2-X_t-\frac{\lambda_tX_t(X_t-1)}{\lambda_t+(\phi-1)X_t}\right]\,\middle|\,\mathcal{F}_{t-1}\right)=\frac{\lambda_t^2}{\phi}+(\phi+1)\lambda_t. \tag{A.15}$$

Using the fact that $X_t^2(X_t-1)=X_t(X_t-1)+X_t(X_t-1)^2$, then from (A.10) and (A.11) we have

$$E\left(\frac{X_t^2(X_t-1)}{[\lambda_t+(\phi-1)X_t]^2}\,\middle|\,\mathcal{F}_{t-1}\right)=\frac{\lambda_t}{\phi^2}+\frac{2\lambda_t}{\phi^2[\lambda_t+2(\phi-1)]}. \tag{A.16}$$

By (A.4), (A.14), (A.15) and (A.16), it is easy to verify that

$$\begin{aligned}E\left(\left[\frac{X_t(X_t-1)}{\lambda_t+(\phi-1)X_t}-\frac{X_t}{\phi}-\frac{X_t-\lambda_t}{\phi^2}\right]^2\,\middle|\,\mathcal{F}_{t-1}\right)&=E\left(\left(\frac{X_t^2(X_t-1)}{[\lambda_t+(\phi-1)X_t]^2}-\frac{X_t}{\phi^2}-\frac{2(X_t-\lambda_t)}{\phi^3}\right)\,\middle|\,\mathcal{F}_{t-1}\right)\\&=\frac{2\lambda_t}{\phi^2[\lambda_t+2(\phi-1)]}.\end{aligned} \tag{A.17}$$

References

[1] T. Amemiya, Advanced Econometrics, Harvard University Press, Cambridge, 1985.
[2] A.C. Cameron, P.K. Trivedi, Regression Analysis of Count Data, Cambridge University Press, New York, 1998.
[3] P.C. Consul, Generalized Poisson Distributions: Properties and Applications, Marcel Dekker, New York, 1989.
[4] P.C. Consul, F. Famoye, Generalized Poisson regression model, Comm. Statist. Theory Methods 21 (1992) 89–109.
[5] P.C. Consul, F. Famoye, Lagrangian Probability Distributions, Birkhäuser, Boston, 2006.
[6] P.C. Consul, G.C. Jain, A generalization of the Poisson distribution, Technometrics 15 (1973) 791–799.
[7] P.C. Consul, M.M. Shoukri, The negative integer moments of the generalized Poisson distribution, Comm. Statist. Simulation Comput. 15 (1986) 1053–1064.
[8] B. Efron, Double exponential families and their use in generalized linear regression, J. Amer. Statist. Assoc. 81 (1986) 709–721.
[9] F. Famoye, Restricted generalized Poisson regression model, Comm. Statist. Theory Methods 22 (1993) 1335–1354.
[10] F. Famoye, J.T. Wulu, K.P. Singh, On the generalized Poisson regression model with an application to accident data, J. Data Sci. 2 (2004) 287–295.

[11] R. Ferland, A. Latour, D. Oraichi, Integer-valued GARCH process, J. Time Ser. Anal. 27 (2006) 923–942.
[12] K. Fokianos, Some recent progress in count time series, Statistics 45 (2011) 49–58.
[13] K. Fokianos, R. Fried, Interventions in INGARCH processes, J. Time Ser. Anal. 31 (2010) 210–225.
[14] K. Fokianos, A. Rahbek, D. Tjøstheim, Poisson autoregression, J. Amer. Statist. Assoc. 104 (2009) 1430–1439.
[15] K. Fokianos, D. Tjøstheim, Log-linear Poisson autoregression, J. Multivariate Anal. 102 (2011) 563–578.
[16] J. Franke, Weak dependence of functional INGARCH processes, Technical report, University of Kaiserslautern, 2010.
[17] A. Heinen, Modeling time series count data: an autoregressive conditional Poisson model, CORE Discussion paper 2003/62, Université catholique de Louvain, 2003.
[18] S.T. Jensen, A. Rahbek, Asymptotic inference for nonstationary GARCH, Econometric Theory 20 (2004) 1203–1226.
[19] R.C. Jung, M. Kukuk, R. Liesenfeld, Time series of count data: Modeling, estimation and diagnostics, Comput. Statist. Data Anal. 51 (2006) 2350–2364.
[20] R.C. Jung, A.R. Tremayne, Useful models for time series of counts or simply wrong ones? Adv. Stat. Anal. 95 (2011) 59–91.
[21] B. Kedem, K. Fokianos, Regression Models for Time Series Analysis, Wiley, New Jersey, 2002.
[22] S. Ling, Estimation and testing stationarity for double autoregressive models, J. Roy. Statist. Soc. Ser. B 66 (2004) 63–78.
[23] D.S. Matteson, M.W. McLean, D.B. Woodard, S.G. Henderson, Forecasting emergency medical service call arrival rates, Ann. Appl. Stat. 5 (2011) 1379–1406.
[24] E. McKenzie, Discrete variate time series, in: D.N. Shanbhag, C.R. Rao (Eds.), Handbook of Statistics, vol. 21, Elsevier Science, Amsterdam, 2003, pp. 573–606.
[25] M.H. Neumann, Absolute regularity and ergodicity of Poisson count processes, Bernoulli 17 (2011) 1268–1284.
[26] İ. Özmen, Quasi likelihood/moment method for generalized and restricted generalized Poisson regression models and its application, Biom. J. 42 (2000) 303–314.
[27] M.S. Ridout, P. Besbeas, An empirical model for underdispersed count data, Stat. Model. 4 (2004) 77–89.
[28] W. Wang, F. Famoye, Modeling household fertility decisions with generalized Poisson regression, J. Population Econ. 10 (1997) 273–283.
[29] C.H. Weiß, Controlling correlated processes of Poisson counts, Qual. Reliab. Eng. Internat. 23 (2007) 741–754.
[30] C.H. Weiß, Thinning operations for modeling time series of counts—a survey, Adv. Stat. Anal. 92 (2008) 319–341.
[31] C.H. Weiß, Serial dependence and regression of Poisson INARMA models, J. Statist. Plann. Inference 138 (2008) 2975–2990.
[32] C.H. Weiß, Modelling time series of counts with overdispersion, Stat. Methods Appl. 18 (2009) 507–519.
[33] C.H. Weiß, The INARCH(1) model for overdispersed time series of counts, Comm. Statist. Simulation Comput. 39 (2010) 1269–1291.
[34] C.H. Weiß, INARCH(1) processes: Higher-order moments and jumps, Statist. Probab. Lett. 80 (2010) 1771–1780.
[35] H. White, Maximum likelihood estimation of misspecified models, Econometrica 50 (1982) 1–25.
[36] F. Zhu, A negative binomial integer-valued GARCH model, J. Time Ser. Anal. 32 (2011) 54–67.
[37] F. Zhu, Zero-inflated Poisson and negative binomial integer-valued GARCH models, J. Statist. Plann. Inference (2012), doi:10.1016/j.jspi.2011.10.002, in press.
[38] F. Zhu, Q. Li, Moment and Bayesian estimation of parameters in the INGARCH $(1,1)$ model, J. Jilin Univ. (Science Edition) 47 (2009) 899–902 (in Chinese).
[39] F. Zhu, Q. Li, D. Wang, A mixture integer-valued ARCH model, J. Statist. Plann. Inference 140 (2010) 2025–2036.
[40] F. Zhu, D. Wang, Diagnostic checking integer-valued ARCH(p) models using conditional residual autocorrelations, Comput. Statist. Data Anal. 54 (2010) 496–508.
[41] F. Zhu, D. Wang, Estimation and testing for a Poisson autoregressive model, Metrika 73 (2011) 211–230.
[42] F. Zhu, D. Wang, F. Li, H. Li, Empirical likelihood inference for an integer-valued ARCH(p) model, J. Jilin Univ. (Science Edition) 46 (2008) 1042–1048 (in Chinese).
[43] W. Zucchini, I.L. MacDonald, Hidden Markov Models for Time Series, CRC Press, Boca Raton, 2009.