Communications for Statistical Applications and Methods
2019, Vol. 26, No. 3, 295–304

https://doi.org/10.29220/CSAM.2019.26.3.295
Print ISSN 2287-7843 / Online ISSN 2383-4757

Threshold-asymmetric volatility models for integer-valued time series

Deok Ryun Kim[a,b], Jae Eun Yoon[a], Sun Young Hwang[1,a]

[a] Department of Statistics, Sookmyung Women's University, Korea;
[b] International Vaccine Institute, Korea

Abstract

This article deals with threshold-asymmetric volatility models for over-dispersed and zero-inflated time series of count data. We introduce various threshold integer-valued autoregressive conditional heteroscedasticity (ARCH) models as incorporating over-dispersion and zero-inflation via conditional Poisson and negative binomial distributions. EM-algorithm is used to estimate parameters. The cholera data from Kolkata in India from 2006 to 2011 is analyzed as a real application. In order to construct the threshold-variable, both local constant mean which is time-varying and grand mean are adopted. It is noted via a data application that threshold model as an asymmetric version is useful in modelling count time series volatility.

Keywords: count data, integer-valued time series, threshold integer-valued ARCH, volatility

1. Introduction

Over the past three decades, there has been increasing interest in modeling integer-valued time series because of the broad range of potential applicability to epidemiology (Cardinal *et al.*, 1999; Yoon and Hwang, 2015a, 2015b), social science (McCabe and Martin, 2005; Truong *et al.*, 2017), experimental biology (Zhou and Basawa, 2005; Bartlett and McCormick, 2017), environmental science (Thyregod *et al.*, 1999; Pavlopoulos and Karlis, 2008), and economics (Freeland and McCabe, 2004; Quoreshi, 2014).

The first order integer-valued autoregressive model (INAR(1)) with Poisson distribution has been introduced by McKenzie (1985). Alzaid and Al-Osh (1990) extended it to p^{th}-order model (INAR(p)) and Al-Osh and Alzaid (1988) introduced a q^{th}-order integer-valued moving average model (INMA(q)). A integer-valued ARMA models for dependent sequences of Poisson counts was investigated by McKenzie (1988).

Although modeling time series of count data with Poisson distribution is a useful tool, in practice, over-dispersion and zero-inflation in the time series is easily led to a violation of major assumptions that the variance is equal to the mean, and the parameters are to be positive. Ferland *et al.* (2006) proposed an integer-valued GARCH model to study over-dispersed counts, and Fokianos and Fried (2010), Weiß (2010), Zhu and Wang (2010), Zhu (2011, 2012a), and Yoon and Hwang (2015a, 2015b) made further contributions to the literature. Zhu (2012b) extended the model to address both over-dispersion and zero inflation phenomenon in count data. Yoon and Hwang (2015b) presented a data application of the zero-inflated model via conditional Poisson and negative binomial marginals. Wang

[1] Corresponding author: Department of Statistics, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Seoul 04310, Korea. E-mail: shwang@sookmyung.ac.kr

Published 31 May 2019 / journal homepage: http://csam.or.kr

et al. (2014) proposed a self-exited threshold integer-valued Poisson autoregression model (SETPAR) which allows for more general modeling framework including the possibility of negative serial dependence in the time series of count data.

In this paper we study conditional variance (volatility) for over-dispersion, zero-inflation, and serial dependence of count time series data. The organization of this paper is as follows. Section 2 re-introduces existing models as threshold integer-valued analogue of the autoregressive conditional herosckedastic (ARCH) model by adding threshold-asymmetric effects to the models. It is noted that innovation follows either Poisson distribution or negative binomial distribution. Their estimation method is discussed in Section 3. Section 4 illustrates appropriate threshold model building strategies via applying proposed threshold-models to actual, highly skewed, zero-inflated, and serially correlated data example of cholera disease in Kolkata in India from 2006 to 2011 (Ali *et al.*, 2016). Concluding remarks are presented in Section 5.

2. Various integer-valued threshold-asymmetric ARCH models

2.1. INTARCH(1) model

The first-order integer-valued ARCH (INARCH(1), for short) model (Ferland, 2006) is defined as a conditional Poisson model defined by

$$X_t|F_{t-1} \sim \text{Poisson}(\lambda_t), \quad \lambda_t = \alpha_0 + \alpha_1 X_{t-1}, \tag{2.1}$$

where F_{t-1} denotes the information available up to time $t-1$. The parameters α_0 and α_1 are positive. The model (2.1) may be extended to have a two-regime structure of the conditional mean process according to the magnitude of the observation. A threshold model based on (2.1), so called, the first-order integer-valued threshold ARCH (INTARCH(1), for short) model is defined as (Wang *et al.*, 2014)

$$\begin{gathered} X_t|F_{t-1} \sim \text{Poisson}(\lambda_t), \quad \lambda_t = \alpha_0 + \alpha_1 X_{t-1}^{(r)} + \alpha_2 X_{t-1}^{(l)}, \\ X_{t-1}^{(r)} = \begin{cases} X_{t-1}, & \text{if } X_{t-1} > m_t, \\ 0, & \text{if } X_{t-1} \le m_t, \end{cases} \quad X_{t-1}^{(l)} = \begin{cases} X_{t-1}, & \text{if } X_{t-1} \le m_t, \\ 0, & \text{if } X_{t-1} > m_t, \end{cases} \end{gathered} \tag{2.2}$$

where the parameters α_0, α_1, and α_2 are positive, and the initial value $X_0 = x_0$ is fixed. Here, m_t is a (time varying) threshold variable that determines the dynamic switching mechanism of the model. The dynamics of the process is governed by a two-regime scheme. Specifically, if $X_{t-1} > m_t$ then we say X_t lies in the upper regime, otherwise, X_t belongs to the lower regime. Various choices of the threshold variable have been used in applications (Wu and Chen, 2007). In the real data example, we employ two threshold variables: one is simply grand mean of the entire time series and the other is the local constant mean which is time-varying.

2.2. NB-INTARCH(1) model

To accommodate over-dispersion in the data, one may consider the model for which negative binomial distribution is used to model the process. The first-order integer-valued negative binomial ARCH (NB-INARCH(1), for short) model (Zhu, 2011; Yoon and Hwang, 2015a) is defined as

$$X_t|F_{t-1} \sim NB(r, p_t), \quad \lambda_t = \frac{1 - p_t}{p_t} = \alpha_0 + \alpha_1 X_{t-1}, \tag{2.3}$$

where the parameter r is a positive integer. We propose the following threshold-asymmetric model as a generalization of (2.3).

$$X_t|F_{t-1} \sim NB(r, p_t), \quad \lambda_t = \frac{1 - p_t}{p_t} = \alpha_0 + \alpha_1 X_{t-1}^{(r)} + \alpha_2 X_{t-1}^{(l)}, \tag{2.4}$$

where $X_{t-1}^{(r)}$ and $X_{t-1}^{(l)}$ are two threshold values defined by (2.2). The model (2.4) is referred to as negative binomial INTARCH(1), denoted as NB-INTARCH(1).

2.3. ZIP-INTARCH(1) model

In order to capture zero-inflation in the count data, Zhu (2012a) and Yoon and Hwang (2015b) investigated the following first-order zero-inflated Poisson ARCH (ZIP-INARCH(1)) model which is formulated as

$$X_t|F_{t-1} \sim \text{ZIP}(\lambda_t, w), \quad \lambda_t = \alpha_0 + \alpha_1 X_{t-1}, \tag{2.5}$$

where $\text{ZIP}(\lambda_t, w)$ is defined as the following probability mass function (pmf)

$$P(X = k) = w\delta_{k,0} + (1 - w)\frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots,$$

where $\delta_{k,0}$ is 1 when $k = 0$ and is zero when $k \neq 0$. The parameter $0 < w < 1$ determines severity of zero-inflation. If $w = 0$, then the model reduces to the Poisson INARCH(1) model defined in (2.1). The ZIP-INARCH(1) model is made to be threshold-asymmetric via

$$X_t|F_{t-1} \sim \text{ZIP}(\lambda_t, w), \quad \lambda_t = \alpha_0 + \alpha_1 X_{t-1}^{(r)} + \alpha_2 X_{t-1}^{(l)}, \tag{2.6}$$

where $X_{t-1}^{(r)}$ and $X_{t-1}^{(l)}$ are two threshold values as defined in (2.2). The model (2.6) is regarded as zero-inflated Poisson INTARCH(1), that is, ZIP-INTARCH(1) model.

2.4. ZINB-INTARCH(1) model

With replacing Poisson by negative binomial distribution in (2.5), the ZIP-INARCH(1) model becomes the following first-order zero-inflated negative binomial INARCH (ZINB-INARCH(1), for short) model which was considered in Zhu (2012a) and Yoon and Hwang (2015b), given as

$$X_t|F_{t-1} \sim \text{ZINB}(\lambda_t, a, w), \quad \lambda_t = \frac{1 - p_t}{p_t} = \alpha_0 + \alpha_1 X_{t-1}, \tag{2.7}$$

where $\text{ZINB}(\lambda_t, a, w)$ is defined by

$$P(X = k) = w\delta_{k,0} + (1 - w)\frac{\Gamma\left(k + \frac{\lambda^{1-c}}{a}\right)}{k!\Gamma\left(\frac{\lambda^{1-c}}{a}\right)}\left(\frac{1}{1 + a\lambda^c}\right)^{\frac{\lambda^{1-c}}{a}}\left(\frac{a\lambda^c}{1 + a\lambda^c}\right)^k, \quad k = 0, 1, 2, \ldots, \tag{2.8}$$

where $0 < w < 1$, $\lambda > 0$, the dispersion parameter $a > 0$, and Γ denotes the standard Gamma function. The index c $(= 0, 1)$ identifies the particular form of the underlying NB distribution (Ridout *et al.*, 2001). For $c = 0$, this particular distribution is denoted by $\text{ZINB1}(\lambda_t, a, w)$ and the case of $c = 1$ refers to ZINB2(λ_t, a, w). The ZINB-INARCH(1) model is now equipped with the threshold-asymmetry by using the equation

$$X_t|F_{t-1} \sim \text{ZINB}(\lambda_t, a, w), \quad \lambda_t = \frac{1-p_t}{p_t} = \alpha_0 + \alpha_1 X_{t-1}^{(r)} + \alpha_2 X_{t-1}^{(l)}. \tag{2.9}$$

Here $X_{t-1}^{(r)}$ and $X_{t-1}^{(l)}$ are again two threshold values described in (2.2). The model (2.9) can be referred to as zero-inflated negative binomial INTARCH(1), abbreviated as ZINB-INTARCH(1).

3. Estimation of parameters

For each model discussed in Section 2, we use EM algorithm to estimate the parameters following the method proposed by Zhu (2012a). See also Yoon and Hwang (2015a, 2015b) for the application of EM algorithm in the context of count time series. Since general steps are the same except conditional log-likelihood function, first-order and second-order derivatives of the log-likelihood function with respect to parameters, we discuss INTARCH(1) case only, defined by (2.2). Let $\theta = (\alpha_0, \alpha_1, \alpha_2)$.

The likelihood function of Poisson distribution in INTARCH(1) model is

$$L(\theta) = \prod_{t=1}^{n} \frac{\lambda_t e^{-\lambda_t}}{X_t}.$$

The conditional log-likelihood function is

$$l(\theta) = \log L(\theta) = \sum_{t=1}^{n} X_t \log \lambda_t - \lambda_t - \log(X_t!).$$

The first derivative of the log-likelihood with respect to $\theta = (\alpha_0, \alpha_1, \alpha_2)$ is given by

$$\frac{\partial l(\theta)}{\partial \theta_i} = \left\{ \frac{X_t}{\lambda_t} - 1 \right\} \frac{\partial \lambda_t}{\partial \theta_i},$$

while the second derivative is

$$\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j} = \left\{ \frac{X_t}{\lambda_t} - 1 \right\} \frac{\partial^2 \lambda_t}{\partial \theta_i \partial \theta_j} - \left\{ \frac{X_t}{\lambda_t^2} \right\} \frac{\partial \lambda_t}{\partial \theta_i} \frac{\partial \lambda_t}{\partial \theta_j}.$$

The iterative EM procedure estimates the parameter $\theta = (\alpha_0, \alpha_1, \alpha_2)$ by maximizing the log-likelihood function. It consists of an E step and M step described as follows:

- **E step**: Given initial values of $\theta^{(0)}$ and X_0, calculate λ_t. And in case of zero-inflation model, the missing data are replaced by their conditional expectation which is given by $w/(w + (1-w)e^{-\lambda_t})$ and zero respectively, according to $X_t = 0$ and $X_t = 1, 2, \ldots$ (Zhu, 2012a, 2012b). In the subsequent iteration, the estimated values in M step are used to calculate λ_t and the conditional expectation of missing data in case of zero-inflation model.

- **M step**: The estimation of θ can be obtained by maximizing the log-likelihood function. Starting with initial values $\theta^{(0)}$, the values of θ in the subsequent iteration can be obtained as

$$\theta^{(i+1)} = \theta^{(i)} - \left\{ \left. \frac{\partial^2 l}{\partial \theta \partial \theta^T} \right|_{\theta^{(i)}} \right\}^{-1} \left. \frac{\partial l}{\partial \theta} \right|_{\theta^{(i)}},$$

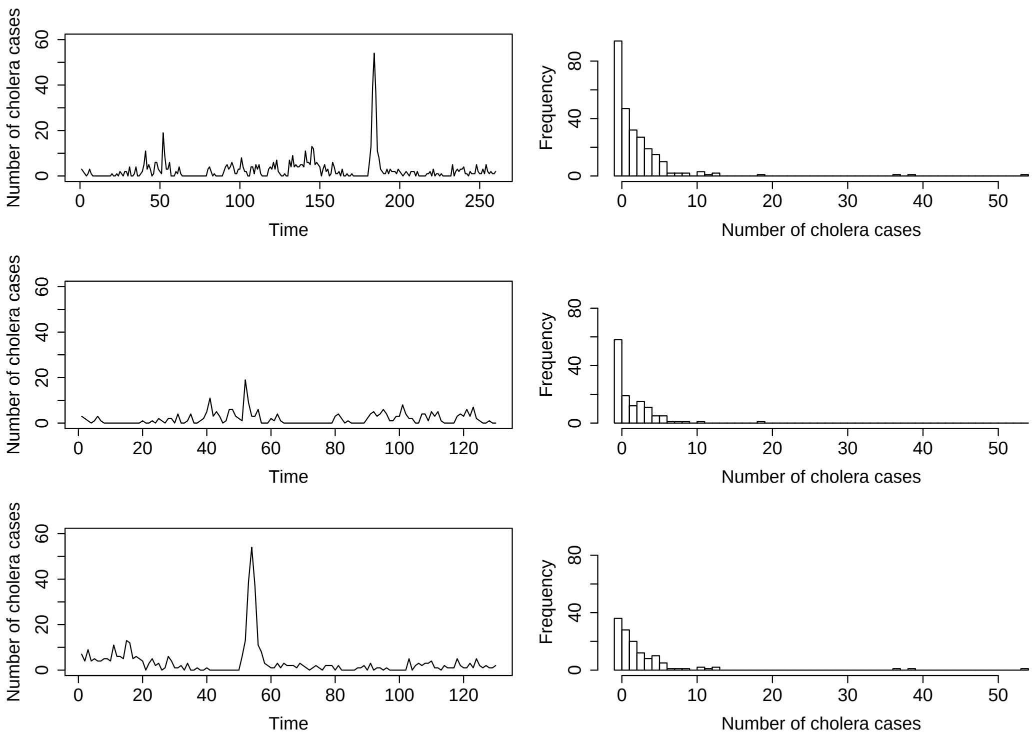

Figure 1: *Weekly cholera cases series and histogram of cholera cases for entire period (n = 260; upper panel), 1^{st} half period (n = 130; middle panel), 2^{nd} half period (n = 130; lower panel).*

where $\theta^{(i)}$ is the value in the i^{th} iteration. The missing data and λ_t are set from the previous E step of the EM procedure.

The estimates $\theta = (\alpha_0, \alpha_1, \alpha_2)$ are obtained by iterating these two steps until convergence. In the data analysis presented in the next section, the convergence criterion of the EM procedure is given as $|(\theta^{(i+1)} - \theta^{(i)})/\theta^{(i)}| \leq 10^{-5}$.

4. Real data analysis: the cholera data from Kolkata in India

In this section, via real data application, we illustrate INTARCH, NB-INTARCH, ZIP-INTARCH, and ZINB-INTARCH models which are defined in Section 3. We consider time series of weekly cholera cases from Kolkata in India, consisting of 260 observations starting from 39^{th}-week of 2006 to 38th-week of 2011. Figure 1 shows the time series plots and histograms for the entire period (in upper panel), 1^{st} half period (in middle panel) and 2^{nd} half period (in lower panel). For entire period, empirical mean and variance of the data are 2.58 and 27.51, respectively.

A histogram of the series shows there are 94 zeros which is 36% of the series. The zero-inflation index (zi) defined by Puig and Valero (2006) to measure the departure from the Poisson model is calculated using the formula

$$zi = 1 + \frac{\log(p_0)}{\mu}, \tag{4.1}$$

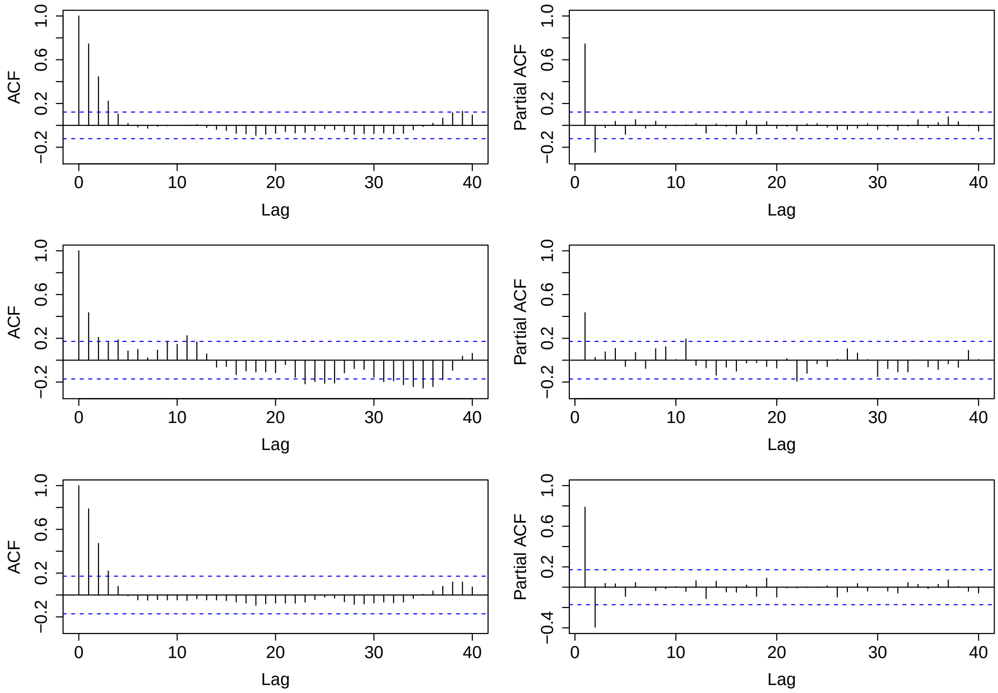

Figure 2: *Sample autocorrelation functions and sample partial autocorrelation functions of the series for entire period (upper panel), 1st half period (middle panel) and 2nd half period (lower panel).*

where p_0 is the proportion of zero's and μ is the mean. It is noted that zi is zero if the count time series data is Poisson-distributed and $zi > 0$ if the count time series data is zero-inflated. The zero-inflation index in the entire period is 0.61, which indicates there is a zero inflation. Similarly, the time series of 1st half period and 2^{nd} half period also have high zero-inflation index, 0.57 and 0.62, respectively. However, when comparing these two period, the 1st half period shows lower mean (1.85) and variance (7.07) but higher proportion of zero counts (45%), while the second half period shows higher mean (3.3) and variance (47.1) but lower proportion of zero counts (28%).

The sample autocorrelation and partial autocorrelation function of the series in entire period, 1^{st} half period, and 2^{nd} half period are plotted in Figure 2 from which it is noted that there exists serial dependency. As in Tables 1–3 below, we consider various models including INTARCH(1), NB-INTARCH(1), ZIP-INTARCH(1) and ZINB-INTARCH(1). Each model is fitted to the entire period and then is fitted further separately to two subset periods: first half and second half period.

The initial values $\theta^{(0)} = (\alpha_0^{(0)}, \alpha_1^{(0)}, \alpha_2^{(0)})$ are randomly selected from the uniform distribution over the unit interval (0,1). Initial values for $w = 0.5$ and $a = 0.1$ are chosen arbitrary and first observation is taken as the initial value for X_0. We then iterate the estimation procedure by using the previous estimated value as the next initial value and we stop the procedure when the convergence criteria $|(\theta^{(i+1)} - \theta^{(i)})/\theta^{(i)}| \leq 10^{-5}$ is met. Motivated by incubation period (Azman *et al.*, 2013) and transmission period (Ali *et al.*, 2016), the most recent 4 time points (which is 4 weeks (28 days)) is considered as local constant mean m_t which is used as a threshold variable. Specifically,

$$m_t = [\text{average of } (X_{t-4} + X_{t-3} + X_{t-2} + X_{t-1}) + 0.5], \tag{4.2}$$

Table 1: Parameter estimates: entire period

Models	Threshold value	$\hat{w}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{a}$	AIC	BIC	$-2\log L(\hat{\theta})$
INARCH(1)			0.79603	0.69163			1129.8	1136.9	1125.8
INTARCH(1)	Grand mean		0.79851	0.69258	0.68577		1131.8	1136.9	1125.8
INTARCH(1)	Local constant		0.76621	0.66631	0.77497		1130.0	1135.2	1124.0
NB1-INARCH(1)			0.73410	0.68364		0.99999	959.0	964.1	953.0
NB1-INTARCH(1)	Grand mean		0.73374	0.68350	0.68450	0.99999	961.0	964.1	953.0
NB1-INTARCH(1)	Local constant		0.70456	0.65637	0.77028	0.99999	959.2	962.3	951.2
NB2-INARCH(1)			0.81757	0.66694		0.69629	957.5	962.6	951.5
NB2-INTARCH(1)	Grand mean		0.80786	0.65285	0.69834	0.69632	959.4	962.5	951.4
NB2-INTARCH(1)	Local constant		0.78486	0.58835	0.80545	0.68720	957.7	960.8	949.7
ZIP-INARCH(1)		0.19694	1.15342	0.69644			1097.1	1102.2	1091.1
ZIP-INTARCH(1)	Grand mean	0.19639	1.13052	0.69351	0.73603		1098.9	1102.0	1090.9
ZIP-INTARCH(1)	Local constant	0.18967	1.07684	0.67218	0.81824		1096.7	1099.8	1088.7
ZNB1-INARCH(1)		0.15709	0.99536	0.70046		0.99999	971.8	974.9	963.8
ZNB1-INTARCH(1)	Grand mean	0.15634	0.96485	0.69485	0.75873	0.99999	973.2	974.3	963.2
ZNB1-INTARCH(1)	Local constant	0.15112	0.92517	0.66899	0.83502	0.99999	970.2	971.3	960.2
ZNB2-INARCH(1)		0.00001	0.81759	0.66694		0.69626	959.5	962.6	951.5
ZNB2-INTARCH(1)	Grand mean	0.00001	0.80787	0.65285	0.69835	0.69630	961.4	962.5	951.4
ZNB2-INTARCH(1)	Local constant	0.00001	0.78487	0.58835	0.80546	0.68717	959.7	960.8	949.7

AIC = Akaike information criterion; BIC = Bayesian information criterion.

where $[x]$ denotes the greatest integer function not exceeding x.

The results of model fitting for entire period and two subset periods are summarized in Tables 1–3, respectively. It is noted that NB is further splitted as NB1 and NB2 according to the index $c = 0$ and $c = 1$, respectively, defined in the pmf equation (2.8). In conclusion, negative binomial models and zero-inflated negative binomial models accommodating over-dispersion and zero-inflation are best fitted both in the entire period data and in each subset period data. Considering Table 1, based on Akaike information criterion (AIC) and Bayesian information criterion (BIC), we find that negative binomial model and zero-inflated negative binomial models are more appropriate. Among the models, ZINB1-INTARCH(1) model shows that a substantial improvement occurs when using the local constant threshold instead of grand mean threshold. To assess the adequacy of the "threshold" ZINB1-INTARCH(1) model over the "non-threshold" ZINB1-INARCH(1) model, we use the following likelihood ratio test (LRT) statistic

$$\text{LRT} = -2\log L(\theta^{*}) - \left[-2\log L(\theta^{\star})\right],$$

where $\log L(\theta^{*})$ and $\log L(\theta^{\star})$ denotes log-likelihood function of ZINB1-INARCH(1) model and log-likelihood of ZINB1-INTARCH(1) model. Note that ZINB1-INARCH(1) is nested in ZINB1-INTARCH(1). Due to Self and Liang (1987), the asymptotic distribution of LRT is given by the $50:50$ mixture of the constant zero and the $\chi^2(1)$ distribution under the null. It is noted that LRT is calculated as 3.6 which is slightly lower than 3.84 which is the upper 5 percent of $\chi^2(1)$ distribution. Consequently, the adequacy of ZINB1-INTARCH(1) over ZINB1-INARCH(1) is highly significant with the p-value given by $P(\text{null distribution exceeds } 3.6) = P(\chi^2(1) > 6.2)$. Another careful examination of Table 1 indicates that NB2-INTARCH(1) and ZINB2-INTARCH(1) using local constant are appropriate models. This confirms that there exist both over-dispersion and zero-inflation in the entire data.

In each subset period (see Tables 2 and 3), it is seen overall that negative binomial models and zero-inflated negative binomial models are appropriate. See the ZINB1-INTARCH(1) model fitted in the 1st half period series and observe that a substantial improvement is obtained when using local

Table 2: Parameter estimates: 1^{st} half period

Models	Threshold value	$\hat{w}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{a}$	AIC	BIC	$-2\log L(\hat{\theta})$
INARCH(1)			0.74651	0.59011			516.7	522.4	512.7
INTARCH(1)	Grand mean		0.65236	0.56924	0.96502		514.6	518.3	508.6
INTARCH(1)	Local constant		0.67205	0.52123	0.84128		513.2	516.9	507.2
NB1-INARCH(1)			0.59222	0.59579		0.99999	431.2	434.9	425.2
NB1-INTARCH(1)	Grand mean		0.57416	0.58583	0.69974	0.99999	432.9	434.6	424.9
NB1-INTARCH(1)	Local constant		0.56399	0.54626	0.74637	0.99999	431.9	433.6	423.9
NB2-INARCH(1)			0.61644	0.76296		0.99596	438.5	442.2	432.5
NB2-INTARCH(1)	Grand mean		0.56781	0.66614	1.16379	0.96519	438.6	440.3	430.6
NB2-INTARCH(1)	Local constant		0.57547	0.59934	1.07206	0.95667	438.1	439.8	430.1
ZIP-INARCH(1)		0.39696	2.15690	0.33124			488.1	491.8	482.1
ZIP-INTARCH(1)	Grand mean	0.36911	1.68879	0.40966	0.90624		488.4	490.1	480.4
ZIP-INTARCH(1)	Local constant	0.36896	1.73160	0.35980	0.69876		486.9	488.6	478.9
ZNB1-INARCH(1)		0.02876	0.61859	0.61160		0.99999	433.1	434.8	425.1
ZNB1-INTARCH(1)	Grand mean	0.03345	0.60215	0.60188	0.73599	0.99999	434.8	434.5	424.8
ZNB1-INTARCH(1)	Local constant	0.05860	0.60291	0.55509	0.87566	0.99999	432.9	432.6	422.9
ZNB2-INARCH(1)		0.00001	0.61646	0.76294		0.99593	440.5	442.2	432.5
ZNB2-INTARCH(1)	Grand mean	0.00002	0.56783	0.66614	1.16382	0.96514	440.6	440.3	430.6
ZNB2-INTARCH(1)	Local constant	0.00003	0.57550	0.59934	1.07207	0.95660	440.1	439.8	430.1

AIC = Akaike information criterion; BIC = Bayesian information criterion.

Table 3: Parameter estimates: 2^{nd} half period

Models	Threshold value	$\hat{w}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{a}$	AIC	BIC	$-2\log L(\hat{\theta})$
INARCH(1)			0.79802	0.75262			585.8	591.5	581.8
INTARCH(1)	Grand mean		1.06586	0.80077	0.26593		571.6	575.3	565.6
INTARCH(1)	Local constant		0.79866	0.75308	0.75105		587.8	591.5	581.8
NB1-INARCH(1)			0.76593	0.74109		0.99999	513.6	517.3	507.6
NB1-INTARCH(1)	Grand mean		1.02377	0.78768	0.28480	0.99999	511.4	513.1	503.4
NB1-INTARCH(1)	Local constant		0.76067	0.73735	0.75371	0.99999	515.5	517.2	507.5
NB2-INARCH(1)			0.87897	0.66871		0.46770	500.5	504.3	494.5
NB2-INTARCH(1)	Grand mean		1.06572	0.79207	0.27063	0.40640	495.7	497.4	487.7
NB2-INTARCH(1)	Local constant		0.86367	0.64001	0.71802	0.46731	502.4	504.1	494.4
ZIP-INARCH(1)		0.11934	0.97824	0.77646			575.9	579.6	569.9
ZIP-INTARCH(1)	Grand mean	0.10699	1.27958	0.80541	0.24493		565.5	567.2	557.5
ZIP-INTARCH(1)	Local constant	0.12072	0.97821	0.77481	0.78243		577.9	579.6	569.9
ZNB1-INARCH(1)		0.10186	0.91877	0.76517		0.99999	519.5	521.2	511.5
ZNB1-INTARCH(1)	Grand mean	0.08183	1.18010	0.79731	0.27344	0.99999	516.6	516.3	506.6
ZNB1-INTARCH(1)	Local constant	0.10163	0.90772	0.75933	0.78765	0.99999	521.3	521.0	511.3
ZNB2-INARCH(1)		0.00001	0.87897	0.66873		0.46768	502.5	504.3	494.5
ZNB2-INTARCH(1)	Grand mean	0.00002	1.06575	0.79208	0.27063	0.40637	497.7	497.4	487.7
ZNB2-INTARCH(1)	Local constant	0.00001	0.86366	0.64004	0.71803	0.46729	504.4	504.1	494.4

AIC = Akaike information criterion; BIC = Bayesian information criterion.

constant threshold rather than using grand mean threshold. However, in the Table 3 (2^{nd} half period), the ZINB2-INTARCH(1) model is improved by using grand mean threshold instead of local constant threshold. This might be partly due to the existence of high peak season in 2^{nd} half period.

5. Concluding remarks

In this paper we have discussed various threshold-asymmetric (ARCH-type conditionally heteroscedastic) volatility models to analyze integer-valued count time series. Over-dispersion and zero-inflation are accommodated using negative binomial distributions. The EM method is adopted to estimate parameters. Two threshold variables, viz., grand mean and local constant mean, are considered in various threshold models. It is noted that the local constant mean works usually better than the grand mean while the grand mean seems better than the local constant mean in case when high peak season is prominent in short time period (see 2^{nd} half period). We have compared models using likelihood-based approaches of AIC, BIC, and log-likelihood. Non-likelihood approaches such as forecasting error evaluations via parametric bootstrap can also be implemented to compare various threshold models and this task is now under investigation.

Acknowledgement

We are grateful for sharing the data in this paper to International Vaccine Institute (IVI), Seoul, Korea and the National Institute of Cholera and Enteric Disease (NICED), Kolkata in India who jointly owned the data. We thank the two anonymous referees for careful reading of the paper. SY Hwang's work was supported by a grant from the National Research Foundation of Korea (NRF-2018R1A2B2004157).

References

Ali M, Debes AK, Luquero FJ, *et al.* (2016). Potential for controlling cholera using a ring vaccination strategy: re-analysis of data from a cluster-randomized clinical trial, *PLoS Medicien*, **13**, e1002120.

Al-Osh MA and Alzaid AA (1988). Integer-valued moving average (INMA) process, *Statistical Papers*, **29**, 281–300.

Alzaid AA and Al-Osh M (1990). An integer-valued pth-order autoregressive structure INAR(p) process, *Journal of Applied Probability*, **27**, 314–324.

Azman AS, Rudolph KE, Cummings DAT, and Lessler J (2013). The incubation period of cholera: a systematic review, *The Journal of Infection*, **66**, 432–438.

Bartlett A and McCormick WP (2017). Estimation for a first-order bifurcating autoregressive process with heavy-tail innovations, *Stochastic Models*, **33**, 210–228.

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

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

Fokianos K and Fried R (2010). Interventions in INGARCH processes, *Journal of Time Series Analysis*, **31**, 210–225.

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

McCabe BPM and Martin GM (2005). Bayesian prediction of low count time series, *International Journal of Forecasting*, **21**, 315–330.

McKenzie E (1985). Some simple models for discrete variate time series, *Water Resources Bulletin*, **21**, 635–644.

McKenzie E (1988). Some ARMA models for dependent sequences of Poisson counts, *Advances in Applied Probability*, **20**, 822–835.

Pavlopoulos H and Karlis D (2008). INAR(1) modeling of overdispersed count series with an environmental application, *Environmetrics*, **190**, 369–393.

Puig P and Valero J (2006). Count data distributions, *Journal of the American Statistical Association*, **101**, 332–340.

Quoreshi AMMS (2014). A long-memory integer-valued time series model, INARFIMA, for financial application, *Quantitative Finance*, **14**, 2225–2235.

Ridout M, Hinde J, and Demetrio CGB (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives, *Biometrics*, **57**, 219–223.

Self SG and Liang KY (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of American Statistical Association*, **82**, 605–610.

Thyregod P, Carstensen J, Madsen H, and Arnbjerg-Nielsen K (1999). Integer valued autoregressive models for tipping bucket rainfall measurements, *Environmetrics*, **10**, 395–411.

Truong BC, Chen CW, and Sriboonchitta S (2017). Hysteretic Poisson INGARCH model for integer-valued time series, *Statistical Modelling*, **17**, 401–422.

Wang C, Liu H, Yao JF, Davis RA, and Li WK (2014). Self-excited threshold Poisson autoregression, *Journal of the American Statistical Association*, **109**, 777–787.

Weiß CH (2010). The INARCH(1) model for overdispersed time series of counts, *Communications in Statistics: Simulation and Computation*, **39**, 1269–1291.

Wu S and Chen R (2007). Threshold variable determination and threshold variable driven switching autoregressive models, *Statistica Sinica*, **17**, 241–264.

Yoon JE and Hwang SY (2015a). Integer-valued GARCH models for count time series: case study, *Korean Journal of Applied Statistics*, **28**, 115–122.

Yoon JE and Hwang SY (2015b). Zero-inflated INGARCH using conditional Poisson and negative binomial: data application, *Korean Journal of Applied Statistics*, **28**, 583–592.

Zhou J and Basawa IV (2005). Least-squared estimation for bifurcation autoregressive processes, *Statistics & Probability Letters*, **74**, 77–88.

Zhu F (2011). A negative binomial integer-valued GARCH model, *Journal of Time Series Analysis*, **32**, 54–67.

Zhu F (2012a). Zero-inflated Poisson and negative binomial integer-valued GARCH models, *Journal of Statistical Planning and Inference*, **142**, 826–839.

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

Zhu F and Wang D (2010). Diagnostic checking integer-valued ARCH(p) models using conditional residual autocorrelations, *Computational Statistics and Data Analysis*, **54**, 496–508.

Received January 9, 2019; Revised February 22, 2019; Accepted March 5, 2019