A GARCH Type Poisson Model for Time Series of Counts with Cyclically Varying Zero Inflation

Isuru Ratnayake, V A Samaranayake
Missouri University of Science and Technology

Abstract

In this study, we introduce a generalization of the zero inflated Poisson process to model time series of count data that exhibit both generalized conditional heteroskedastic volatility and cyclical behavior in the zero inflation factor. This is a generalization of the zero-inflated Poisson-GARCH model proposed by Fukang Zhu in 2012, which in turn can be considered a generalization of the Autoregressive Conditional Poisson model proposed by Andreas Heinen in 2003. While Heinen's Autoregressive Conditional Poisson model accommodates GARCH type behavior, Zhu introduces zero inflation. Our proposed model goes one step further by incorporating the flexibility to allow the zero inflation parameter to vary cyclically or be driven by an exogenous variable. A method for estimating the proposed model is introduced and its performance studied using Monte-Carlo methods.

Key Words: Integer Valued, Discrete Time Series, Generalized Conditional heteroskedasticity, Periodicity

1. Introduction

The Poisson distribution is used for modeling the number of events occurring in a fixed time interval. However in many real world situations zero counts occur at a higher frequency than one would expect under a regular Poisson process. In order to model these real applications, a zero inflated Poisson model can be considered. The zero-inflated Poisson model considers two mechanisms for generating zeros. One is a binary process where zeros are generated with a certain probability. This is the zero inflation component of the process. If the binary process indicates a non-zero, the actual count is determined by a Poisson process, which in turn can generate additional zeros. Overdispersion can result due to these excessive (Yang et al. 2009), but this may be an added advantage in many situations. The proposed model allows the parameter of the binary mechanism that induces zero inflation to vary either cyclically or be driven by an exogenous variable. In addition, it incorporates a conditional Poisson model where the intensity parameter is dependent on the past.

The Autoregressive Conditional Poisson model was first proposed by Andreas Heinen (2003) and the Integer Valued GARCH Process was proposed by Ferland et al. (2006). They utilized the formulation of the classical generalized autoregressive conditional heteroskedastic (GARCH) (p, q) model to describe how the intensity parameter in a Poisson process would propagate over time, and addressed the problem of maximum likelihood estimation of the parameters. Combining such models with a zero inflation component is obviously a natural step.

Zhu (2011) discussed the modelling of integer valued time series with overdispersion and potential extreme observations. These integer GARCH models are known by their acronym INGARCH. Zhu (2012) stated that the integer INGARCH model is a popular tool for modelling time series of counts and further mentioned that the negative binomial models can also deal with overdispertion. A negative binomial INGARCH model, which is a generalization of the Poisson INGARCH model, was proposed and the stationary conditions were given as well, as the autocorrelation function by Ye et al. (2012). These authors also allowed the negative binomial INGRACH (NB-INGARCH) model to incorporate covariates, so that the relationship between a time series of counts and correlated external factors can be properly modeled. Further, Zhu (2012) extended the previous study and introduced the zero inflated Poisson integer-valued GARCH model and showed how the EM algorithm can be used to estimate the parameters of the model. The underlying processes in Zhu's models are based on either zero inflated Poisson or zero-inflated negative binomial but do not allow such inflation to be cyclical or governed by any external factor.

In this paper, we try to introduce a generalization of the zero inflated Poisson Process for modeling time series of count data that exhibit both zero inflation and autoregressive conditional Poisson type behaviour while at the same time have the flexibility to allow the zero inflation parameter to vary cyclically or be driven by exogenous set of variables.

2. The Zero Inflated Poisson (ZIP) INGARCH Model with Cyclically Varying Zero Inflation

As given in Zhu (2012), probability mass function (pmf) of ZIP (λ, ω) can be written in the form

$$\Pr(X = k) = \omega\delta_{k,0} + (1-\omega)\frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0,1,2, \dots, \text{ where } 0 < \omega < 1,$$
$$\delta_{k,0} = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0. \end{cases}$$

Again, from Zhu (2012), the mean and the variance of the ZIP distribution are as follows:

$$E(X) = (1-\ \omega)\ \lambda\ \ , var(X) = (1-\ \omega)\lambda(1+\omega\lambda) > E(X).$$

Let us consider the time series of counts $\{X_t\}$. Assume that, conditional on F_{t-1}, the random variables $X_1X_2, \dots, X_n$ are independent, and the conditional distribution of X_t is specified by a ZIP (λ_t, ω_t)distribution. Here ω_t is the probability that the observation is zero and λ_t is the mean of the Poisson process at time t. To be specific, we consider the following model:

$$X_t|F_{t-1} \sim ZIP(\lambda_t, \omega_t) \text{ , where } \omega_t = g(S, \underline{\Gamma})$$

$$\lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \tag{2.1}$$

where $0 < \omega_t < 1\ \forall t$ and $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, i = 1,.,p, j = 1,.,q, p \geq 1, q \geq 0$. F_{t-1} is the $\sigma-$field generated by $\{X_{t-1}, X_{t-2,} \dots\}$, $\omega_t = g(S, \underline{\Gamma})$ is the cyclically varying zero inflation function. Note that S is the duration of a cycle and $\underline{\Gamma}$ is the vector of parameters. The above model is denoted by ZIP-INGARCH(p,q) with cyclically varying zero inflation. The conditional mean and conditional variance of X_t are given by

$$E(X_t|\mathrm{F}_{t-1}) = (1-\omega_t)\lambda_t\ ,\ Var(X_t|\mathrm{F}_{t-1}) = (1-\omega_t)\ \lambda_t(1+\omega_t\lambda_t) \qquad (2.2)$$

Then,

$$Var(X_t|\mathrm{F}_{t-1}) = (1-\omega_t)\lambda_t(1+\omega_t\lambda_t) > E(X_t|\mathrm{F}_{t-1}) = (1-\omega_t)\lambda_t.$$

Furthermore, using arguments similar to those used by Zhu (2012),

$$\begin{aligned} Var(X_t) &= E\big(Var(X_t|\mathrm{F}_{t-1})\big) + Var\big(E(X_t|\mathrm{F}_{t-1})\big) \\ &= E[(1-\omega_t)\lambda_t(1+\omega_t\lambda_t)] + Var[(1-\omega_t)\lambda_t] \\ &= (1-\omega_t)E[\lambda_t] + (1-\omega_t)\omega_t E[\lambda_t^2] + Var[(1-\omega_t)\lambda_t] \\ &= (1-\omega_t)E[\lambda_t] + (1-\omega_t)\omega_t Var[\lambda_t] + (1-\omega_t)\omega_t(E[\lambda_t])^2 + \\ &\quad Var[(1-\omega_t)\lambda_t] \\ &= Var\ (X_t) > (1-\omega_t)E[\lambda_t] = E[X_t]\ . \end{aligned} \qquad (2.3)$$

The above result (2.3) indicates that ZIP-INGARCH(p,q) with cyclically varying zero inflation can be used to model integer valued time series with overdispersion.

In this study we consider two different cases.

2.1 Case 1: Sinusoidal zero inflated Function

In this case cyclically varying zero inflation function $\omega_t = g(S,\underline{\Gamma})$ is a sinusoidal function of time stated as follows:

$$\omega_t = g\big(S,\underline{\Gamma}\big) = A(\ \sin\left(\tfrac{2\pi}{12} * t\right) + 1) = A(\ \sin\left(\tfrac{\pi}{6} * t\right) + 1), \qquad (2.4)$$

where S is the seasonal length and $\underline{\Gamma} = A$; $0 < A < 0.5$.

2.2 Case 2: Zero inflated function is driven by an exogenous variable

In this study we also accommodate inflation to be driven by exogenous variable and in this case $\omega_t = g(S,\underline{\Gamma})$ is a logistic regression function of some exogenous variable V_t, which follows a purely seasonal autoregressive$(AR(S))$ time series model. Thus,

$$V_t = \eta V_{t-s} + \varepsilon_t;\ \text{where } \varepsilon_t \text{ is } i.i.d. \sim N(0,1);$$

$$\omega_t = g\big(S,\underline{\Gamma}\big) = \frac{1}{1+e^{(-\delta_0-\delta_1 v_t)}}\,, \qquad (2.5)$$

where S is the seasonal length as before and $\underline{\Gamma} = \begin{bmatrix}\delta_0\\ \delta_1\end{bmatrix}_{2*1}$.

3. Estimation

In this section EM algorithm and Maximum Likelihood (MLE) is used to estimates the parameters. The EM algorithm was used in the cyclically varying model while MLE was used in the exogenous variable case.

3.1 The ZIP-INGARCH (p, q) model with sinusoidal zero inflation function

Let $X_1, X_2, \dots, X_n$ be generated from the model (2.1) and let $\{Z_t\}$ be a binary process such that $Z_t \sim Bernoulli(\omega_t)$. Also let $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_n)$,
$\boldsymbol{\theta} = (\alpha_0, \alpha_1, .., \alpha_p, \beta_1, .., \beta_q)^{\boldsymbol{T}} = (\theta_0, \theta_1, \dots, \theta_{p+q})$, $\phi = (\boldsymbol{\omega_t}, \boldsymbol{\theta^T})^{\boldsymbol{T}}$**.**

By emulating the proof of Zhu (2012), the conditional log likelihood can be written as,

$$\boldsymbol{l}(\phi) = \sum_{t=p+1}^{n}\{Z_t.\log(\omega_t) + (1 - Z_t)[\log(1 - \omega_t) + X_t \log(\lambda_t) - \lambda_t - \log(X_t!)]\},$$

$$\text{where } \boldsymbol{\omega_t} = \boldsymbol{g}(\boldsymbol{S}, \underline{\boldsymbol{\Gamma}}) = \boldsymbol{A}\left(\mathbf{sin}\left(\frac{\boldsymbol{\pi}}{\mathbf{6}} * \boldsymbol{t}\right) + \mathbf{1}\right). \tag{3.1}$$

The first derivatives of the log-likelihood with respect to A are given by:

$$\frac{dl}{dA} = \sum_{t=p+1}^{n}\left[\frac{Z_t}{A} - \frac{(1-Z_t)C_t}{1-AC_t}\right] \text{ ; Here } C_t = \sin\left(\frac{\pi}{6} * t\right) + 1 \tag{3.2}$$

$$\frac{dl}{d\theta_i} = \sum_{t=p+1}^{n}(1 - Z_t)\left(\frac{X_t}{\lambda_t} - 1\right)\frac{d\lambda_t}{d\theta_i}, i = 0,1, \dots, p + q. \tag{3.3}$$

EM algorithm is used to estimate the parameters by maximizing 3.1 in the same manner as done by Zhu (2012).
E Step: If ϕ is known. The missing data Z_t are replaced by their expectations, conditional on parameter θ and on the observed data X, which are denoted by τ_t, Then;

$$\tau_t = \begin{cases} \dfrac{AC_t}{AC_t + (1 - AC_t)e^{-\lambda_t}}, & if\ X_t = 0 \\ 0, & if\ X_t \neq 1,2, \dots \end{cases}$$

M Step: Assume that the missing data are known. The estimates of ϕ can be obtained by maximizing (3.1). Then;

$$\hat{A} = \frac{\sum_{t=p+1}^{n} \tau_t}{\sum_{t=p+1}^{n}\left(\sin\left(\frac{\pi t}{6}\right) + 1\right)} \tag{3.4}$$

$$\sum_{t=p+1}^{n}(1 - \tau_t)\left(\frac{X_t}{\lambda_t} - 1\right)\frac{d\lambda_t}{d\theta_i}\Big|\hat{\theta} = 0 \quad i = 0,1, \dots, p + q \tag{3.5}$$

Since there is no closed form solution for equation (3.5) Newton Raphson algorithm is used to obtain estimates. The estimates of ϕ are obtained by iterating E steps and M steps until convergence.

3.2 The ZIP-INGARCH (p, q) model where the zero inflated function is driven by an exogenous variable

The likelihood function for the ZIP-INGARCH model (2.1) is

$$\prod_{X_t=0}\left[\omega_t + (1-\omega_t)e^{-\lambda_t}\right] * \prod_{X_t>0}\left[(1-\omega_t)\frac{\lambda_t^{X_t}e^{-\lambda_t}}{X_t!}\right]. \qquad (3.6)$$

Then the log-likelihood function is

$$\sum_{X_t=0}\log\left[\omega_t + (1-\omega_t)e^{-\lambda_t}\right] + \sum_{X_t>0}\left[\log(1-\omega_t) + X_t log\lambda_t - \lambda_t - \log(X_t!)\right] \qquad (3.7)$$

By maximizing the log likelihood function we can estimate the parameters of the ZIP-INGARCH model (2.5).

4. Simulation

Finite sample performance of the estimators was evaluated using a simulation study. **Matlab** software was utilized to carry out the simulation, and the relevant data were generated by the function **poissrnd**. Length of the time series studies was set to $n = 120$ and also $n = 360$. Three thousand (3,000) simulations runs were carried out for each parameter sample size combination. Parameters combinations used in the simulation can be inferred from Tables 1 – 3 for the sinusoidal case and Table 4 for the exogenous variable case. The profile likelihood functions (3.1) and (3.7) were maximized using the constrained nonlinear optimization function **fmincon** in **Matlab**. The zero inflation $(= \omega_t = g(S, \underline{\Gamma}))$ was used to vary cyclically or be driven by an exogenous set of variable. As was done by Zhu (2012), the Mean Absolute Deviation Error (MADE) was utilized as the evaluation criterion. The MADE is defined as, $\frac{1}{m}\sum_{i=1}^{m}|\widehat{\Phi}_j - \Phi_j|$ where m is the number of replications. Samples of simulated data are given in Figures 1 through 4. Simulation results are reported in Tables 1 through 4.

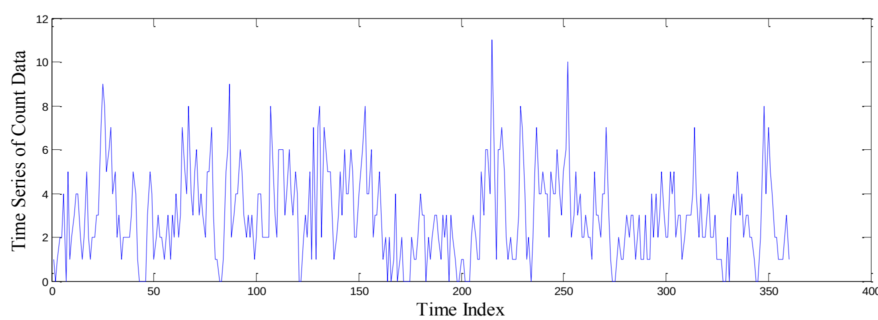

Figure 1: Sample of simulated data for INGARCH (1, 1) model with $A = 0.01$ and $n = 360$

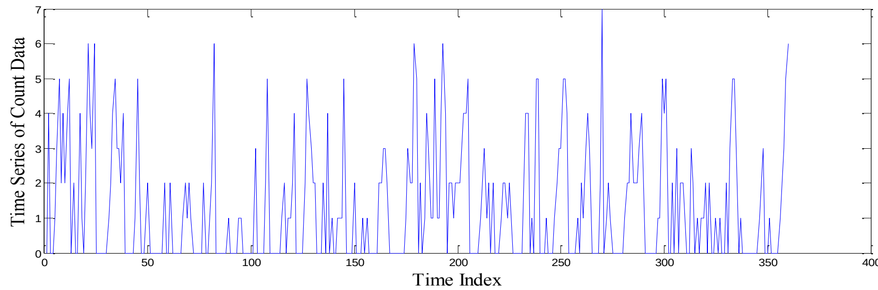

Figure2: Sample of simulated data for INGARCH (1, 1) model with $A = 0.4$ and $n = 360$

It can be clearly seen that the more number of zeros were generated whenever the parameter A increases.

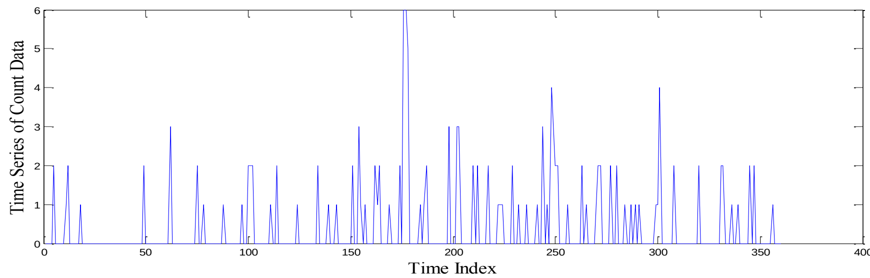

Figure 3: Sample of simulated data for INGARCH (1, 1) model with η= 0.90 and $n = 360$

4.1 Simulation results for Case 1: Sinusoidal zero inflation function

Tables 1 through 3 provide the simulation results for the case where the zero inflation varied in s sinusoidal fashion. The frequency of the sinusoidal wave was set at 12, mimicking a 12 month cycle in monthly data. The parameter *A* was set at 0.01, 0.1, and 0.4, representing, minimal, moderate, and large zero inflation.

In case of minimal zero inflation, the parameters of the GARCH portion of the model were estimated quite accurately, for both samples sizes and for all parameter combinations. When the parameter *A* was set to 0.1, indicating moderate zero inflation, the parameter estimates were reasonable, especially the estimate of *A*, but the other parameter estimates showed some bias. This decline in the estimate accuracy magnified when the zero inflation was high (*A*=0.4), but the parameter *A* was estimated very accurately. Note that the simulation study of Zhu did not consider the high zero inflation scenarios we considered. The probability of getting a zero value in the binary component of the model was set to 0.1 by Zhu throughout their study.

Table 1: Means of Estimates and MADE (within parentheses), for sinusoidal zero inflated parameter INGARCH (1, 1) models and $A = 0.01$

Sample Size(n)	True Value A	α_0	α_1	β_1	Estimated Value (MADE) $\hat{A}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
120	0.01	1	0.4	0.2	0.0148	1.0629	0.3924	0.1812
					(0.0147)	(0.1915)	(0.0667)	(0.0804)
360	0.01	1	0.4	0.2	0.0113	1.0264	0.4003	0.1874
					(0.0094)	(0.1435)	(0.0408)	(0.0688)
120	0.01	1	0.2	0.5	0.0172	1.1912	0.2090	0.4294
					(0.0146)	(0.3526)	(0.0600)	(0.1542)
360	0.01	1	0.2	0.5	0.0161	1.1421	0.2183	0.4369
					(0.0108)	(0.3062)	(0.0433)	(0.1375)
120	0.01	1	0.2	0.2	0.0217	1.0522	0.1952	0.1804
					(0.0212)	(0.1649)	(0.0610)	(0.0911)
360	0.01	1	0.2	0.2	0.0154	1.0290	0.1993	0.1848
					(0.0138)	(0.1335)	(0.0420)	(0.0847)
120	0.01	2	0.4	0.2	0.0107	2.1070	0.3938	0.1816
					(0.0090)	(0.3683)	(0.0663)	(0.0798)
360	0.01	2	0.4	0.2	0.0103	2.0431	0.4066	0.1818
					(0.0056)	(0.2691)	(0.0398)	(0.0674)
120	0.01	2	0.2	0.5	0.0110	2.2857	0.2215	0.4270
					(0.0085)	(0.6220)	(0.0605)	(0.1505)
360	0.01	2	0.2	0.5	0.0105	2.2006	0.2376	0.4271
					(0.0052)	(0.5126)	(0.0499)	(0.1302)
120	0.01	2	0.2	0.2	0.0133	2.0584	0.1947	0.1873
					(0.0116)	(0.3043)	(0.0593)	(0.0905)
360	0.01	2	0.2	0.2	0.0110	2.0302	0.2004	0.1889
					(0.0076)	(0.2586)	(0.0403)	(0.0834)
120	0.01	5	0.4	0.2	0.0100	5.1346	0.4120	0.1749
					(0.0073)	(0.7822)	(0.0641)	(0.0781)
360	0.01	5	0.4	0.2	0.0102	5.0140	0.4270	0.1700
					(0.0043)	(0.6013)	(0.0439)	(0.0670)
120	0.01	5	0.2	0.5	0.0100	5.3915	0.2528	0.4190
					(0.0074)	(1.2012)	(0.0735)	(0.1334)
360	0.01	5	0.2	0.5	0.0100	5.0975	0.2756	0.4131
					(0.0042)	(0.9028)	(0.0785)	(0.1112)
120	0.01	5	0.2	0.2	0.0099	5.0827	0.2021	0.1856
					(0.0072)	(0.7118)	(0.0583)	(0.0882)
360	0.01	5	0.2	0.2	0.0102	5.0062	0.2127	0.1846
					(0.0044)	(0.6107)	(0.0396)	(0.0807)

Table 2: Means of Estimates and MADE (within parentheses), for sinusoidal zero inflated parameter INGARCH (1, 1) models and $A = 0.1$

Sample Size(n)	A	α_0	α_1	β_1	$\hat{A}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
120	0.1	1	0.4	0.2	0.1028	0.9947	0.4194	0.1652
					(0.0343)	(0.1687)	(0.0759)	(0.0831)
360	0.1	1	0.4	0.2	0.1031	0.9936	0.4345	0.1545
					(0.0194)	(0.1187)	(0.0537)	(0.0741)
120	0.1	1	0.2	0.5	0.1157	0.9817	0.2667	0.4075
					(0.0343)	(0.2276)	(0.0776)	(0.1259)
360	0.1	1	0.2	0.5	0.1147	0.9535	0.2861	0.3976
					(0.0216)	(0.1724)	(0.0871)	(0.1146)
120	0.1	1	0.2	0.2	0.1013	0.9939	0.2090	0.1777
					(0.0403)	(0.1553)	(0.0630)	(0.0909)
360	0.1	1	0.2	0.2	0.1038	0.9780	0.2218	0.1783
					(0.0236)	(0.1253)	(0.0470)	(0.0832)
120	0.1	2	0.4	0.2	0.1026	1.9198	0.4544	0.1523
					(0.0241)	(0.2951)	(0.0790)	(0.0813)
360	0.1	2	0.4	0.2	0.1021	1.9374	0.4347	0.1380
					(0.0139)	(0.1909)	(0.0693)	(0.0765)
120	0.1	2	0.2	0.5	0.1076	1.8090	0.2943	0.4053
					(0.0242)	(0.3930)	(0.0949)	(0.1125)
360	0.1	2	0.2	0.5	0.1086	1.7368	0.2997	0.4110
					(0.0156)	(0.3441)	(0.0997)	(0.0925)
120	0.1	2	0.2	0.2	0.1019	1.9383	0.2254	0.1789
					(0.0263)	(0.2838)	(0.0632)	(0.0885)
360	0.1	2	0.2	0.2	0.1017	1.9100	0.2396	0.1750
					(0.0148)	(0.2301)	(0.0525)	(0.0811)
120	0.1	5	0.4	0.2	0.1001	4.5829	0.5073	0.1233
					(0.0211)	(0.5818)	(0.1094)	(0.0854)
360	0.1	5	0.4	0.2	0.1005	4.637	0.5168	0.1096
					(0.0124)	(0.4128)	(0.1168)	(0.0910)
120	0.1	5	0.2	0.5	0.1009	4.1472	0.2999	0.4194
					(0.0214)	(1.0011)	(0.0999)	(0.0893)
360	0.1	5	0.2	0.5	0.1002	3.9018	0.3000	0.4371
					(0.0123)	(1.1166)	(0.1000)	(0.0673)
120	0.1	5	0.2	0.2	0.1013	4.6403	0.2630	0.1696
					(0.0217)	(0.6020)	(0.0723)	(0.0832)
360	0.1	5	0.2	0.2	0.1003	4.6896	0.2795	0.1486
					(0.0125)	(0.4269)	(0.0800)	(0.0772)

Table 3: Means of Estimates and MADE (within parentheses), for sinusoidal zero inflated parameter INGARCH (1, 1) models and $A = 0.4$

Sample		True Value			Estimated Value (MADE)			
Size(n)	A	α_0	α_1	β_1	$\hat{A}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
120	0.4	1	0.4	0.2	0.4040 (0.0388)	0.9332 (0.1759)	0.4338 (0.0999)	0.1700 (0.0861)
360	0.4	1	0.4	0.2	0.4507 (0.0226)	0.9330 (0.1231)	0.4511 (0.0736)	0.1585 (0.0805)
120	0.4	1	0.2	0.5	0.4204 (0.0388)	0.8314 (0.2560)	0.2745 (0.0854)	0.4352 (0.1381)
360	0.4	1	0.2	0.5	0.4210 (0.0273)	0.7978 (0.2398)	0.2929 (0.0936)	0.4266 (0.1062)
120	0.4	1	0.2	0.2	0.4050 (0.0407)	0.9316 (0.1686)	0.2164 (0.0731)	0.1890 (0.0931)
360	0.4	1	0.2	0.2	0.4062 (0.0233)	0.9187 (0.1349)	0.2345 (0.0603)	0.1872 (0.0866)
120	0.4	2	0.4	0.2	0.4023 (0.0321)	1.8112 (0.2929)	0.4707 (0.0983)	0.1558 (0.0848)
360	0.4	2	0.4	0.2	0.4029 (0.0182)	1.8213 (0.2152)	0.4919 (0.0954)	0.1380 (0.0802)
120	0.4	2	0.2	0.5	0.4105 (0.0312)	1.5307 (0.5276)	0.2979 (0.0980)	0.4311 (0.1120)
360	0.4	2	0.2	0.5	0.4111 (0.0197)	1.4628 (0.5481)	0.3000 (0.1000)	0.4421 (0.0785)
120	0.4	2	0.2	0.2	0.4021 (0.0315)	1.7941 (0.3096)	0.2428 (0.0731)	0.1907 (0.0900)
360	0.4	2	0.2	0.2	0.4032 (0.0187)	1.7881 (0.2560)	0.2637 (0.0703)	0.1797 (0.0816)
120	0.4	5	0.4	0.2	0.4005 (0.0299)	4.4100 (0.6361)	0.5239 (0.1258)	0.1293 (0.0855)
360	0.4	5	0.4	0.2	0.4002 (0.0169)	4.4922 (0.5123)	0.5366 (0.1366)	0.1093 (0.0920)
120	0.4	5	0.2	0.5	0.4024 (0.0292)	3.5185 (1.4834)	0.3000 (0.1000)	0.4516 (0.0749)
360	0.4	5	0.2	0.5	0.4009 (0.0169)	3.4089 (1.5911)	0.3000 (0.1000)	0.4584 (0.0565)
120	0.4	5	0.2	0.2	0.4010 (0.0289)	4.3594 (0.7155)	0.2790 (0.0828)	0.1808 (0.0834)
360	0.4	5	0.2	0.2	0.4006 (0.0168)	4.3913 (0.6203)	0.2913 (0.0932)	0.1656 (0.0720)

4.2 Simulation study for case 2: Zero inflation function is driven by exogenous variable

In this study, the exogenous variable was allowed to generate zeros through a logistic model as described in Equation (2.5). The parameters in the logistic model δ_0 and δ_1 were set to values -3 and 2 for δ_0 and -2, 0, and 1 for δ_1. Note that larger values result in higher zero inflation. Results show that the estimates of the GARCH parameters are biased towards the lower values, but the estimates for δ_0 and δ_1 are accurate for all parameter and sample size combinations.

Table 4: Means of Estimates and MADE (within parentheses), for INGARCH (1, 1) models where zero inflation is driven by exogenous variable

Sample size (n)	True Value δ_0	δ_1	α_0	α_1	β_1	Estimated Value (MADE) $\widehat{\delta_0}$	$\widehat{\delta_1}$	$\widehat{\alpha_0}$	$\widehat{\alpha_1}$	$\widehat{\beta_1}$
$\eta = .25$										
120	2	0	1	0.4	0.3	2.0357 (0.3005)	-0.0023 (0.1644)	0.9016 (0.1780)	0.3257 (0.1242)	0.2389 (0.1024)
360	2	0	1	0.4	0.3	2.0812 (0.2636)	0.0000 (0.2636)	0.8970 (0.1779)	0.3230 (0.1251)	0.2374 (0.1029)
120	2	1	1	0.4	0.3	2.0660 (0.4020)	1.0357 (0.2890)	0.9005 (0.1776)	0.3247 (0.1255)	0.2382 (0.1040)
360	2	1	1	0.4	0.3	2.0256 (0.3012)	1.0291 (0.1921)	0.9036 (0.1758)	0.3235 (0.1253)	0.2366 (0.1037)
120	2	1	1	0.3	0.3	2.0594 (0.2881)	1.0398 (0.2881)	0.8999 (0.1823)	0.2461 (0.1143)	0.2461 (0.1143)
360	2	1	1	0.3	0.3	2.0173 (0.3160)	1.0298 (0.1958)	0.8996 (0.1770)	0.2454 (0.1125)	0.2454 (0.1125)
120	-3	1	2	0.3	0.4	-3.4852 (0.7600)	1.0127 (0.3423)	1.9035 (0.1742)	0.2384 (0.1015)	0.3256 (0.1229)
360	-3	1	2	0.3	0.4	-3.4679 (0.7047)	1.0199 (0.2804)	1.9024 (0.1784)	0.2380 (0.1015)	0.3247 (0.1238)
120	3	0	1	0.5	0.4	3.2883 (0.5551)	-0.0038 (0.3297)	0.9020 (0.1783)	0.3864 (0.1160)	0.2974 (0.1026)
360	3	0	1	0.5	0.4	3.1748 (0.3693)	0.0014 (0.2425)	0.9033 (0.1716)	0.3899 (0.1121)	0.3004 (0.0996)
120	-3	-2	1	0.5	0.4	-3.1550 (0.9625)	-1.9478 (0.5986)	0.9037 (0.1783)	0.3847 (0.1150)	0.2976 (0.1024)
360	-3	-2	1	0.5	0.4	-3.1042 (0.9321)	-1.9037 (0.5294)	0.9006 (0.1777)	0.3841 (0.1179)	0.2951 (0.1049)
$\eta = 0.90$										
120	2	0	1	0.4	0.3	2.0945 (0.3973)	-0.0010 (0.1746)	0.8970 (0.1779)	0.3221 (0.1242)	0.2365 (0.1021)
360	2	0	1	0.4	0.3	2.0357 (0.3013)	-0.0011 (0.0844)	0.9016 (0.1780)	0.3237 (0.1222)	0.2369 (0.1004)
120	2	1	1	0.4	0.3	2.0744 (0.4215)	1.0488 (0.2596)	0.9025 (0.1768)	0.3243 (0.1249)	0.2374 (0.1031)
360	2	1	1	0.4	0.3	2.0092 (0.3356)	1.0360 (0.1921)	0.9001 (0.1791)	0.3222 (0.1238)	0.2360 (0.1017)
120	2	1	1	0.3	0.3	2.0656 (0.4411)	1.0558 (0.2625)	0.9052 (0.1771)	0.2490 (0.1113)	0.2490 (0.1113)
360	2	1	1	0.3	0.3	1.9973 (0.3286)	1.0360 (0.1972)	0.9009 (0.1796)	0.2474 (0.1131)	0.2474 (0.1131)
120	-3	1	2	0.3	0.4	-3.4007 (0.8053)	1.0386 (0.2727)	1.8915 (0.1817)	0.2346 (0.1034)	0.3198 (0.1256)
360	-3	1	2	0.3	0.4	-3.3078 (0.7399)	1.0218 (0.2128)	1.9001 (0.1805)	0.2386 (0.1003)	0.3256 (0.1222)
120	3	0	1	0.5	0.4	3.3286 (0.5869)	0.0039 (0.2488)	0.9020 (0.1783)	0.3843 (0.1173)	0.2954 (0.1046)
360	3	0	1	0.5	0.4	3.1822 (0.3786)	0.0004 (0.1272)	0.9033 (0.1716)	0.3821 (0.1184)	0.2935 (0.1065)
120	-3	-2	1	0.5	0.4	-3.0822 (1.0605)	-1.9617 (0.5567)	0.8975 (0.1784)	0.3843 (0.1183)	0.2954 (0.1046)
360	-3	-2	1	0.5	0.4	-2.9669 (1.0405)	-1.8792 (0.5001)	0.9024 (0.1798)	0.3864 (0.1164)	0.2941 (0.1029)

5. Conclusions

The study indicates that the parameters of the proposed model with cyclically varying zero-inflation can be estimated by reasonable accuracy by using EM algorithm in cases where the zero inflation is moderate or low. In cases where the zero inflation is influenced by a exogenous process, MLE method can produce fair estimates of the parameter, especially those associated with the zero inflation mechanism.

Acknowledgements

We would like to thank Dr. Fukang Zhu for providing us with his Mathlab code developed to model non seasonal zero inflated Poisson process.

References

Ferland, R., Latour, A., Oraichi, D., 2006. Journal of Time Series Analysis Vol. 27, No.6.

Heinen, A., 2003. Modelling time series count data: an autoregressive conditional Poisson model.

Yang, Z., Hardin, J., Addy, C., 2009. Testing overdispersion in the zero-inflated Poisson model. Journal of Statistical Planning and inference 139, 3340-3353.

Ye, F., Garcia, P., Pourahmadi, M., Lord, D., 2012. Extension of a Negative Binomial GARCH Model: Analyzing the Effects of Gasoline Price and VMT on DUI Fatal Crashes in Texas. Journal of Transportation Research Board No.2279, pp 31-39.

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

Zhu, F. 2012. Modeling time series of counts with COM-Poisson INGARCH models. Mathematical and Computer Modeling Vol. 56,pp 191-240.

Zhu, F. 2012. Zero-inflated Poisson and negative binomial integer-valued GARCH models. Journal of Statistical Planning and inference 142, 826-839.