Moment Properties and Quadratic Estimating Functions for Integer-valued Time Series Models

Nurul Najihah Mohamad
Department of Computational and Theoretical Sciences,
Kulliyyah of Science, International Islamic University Malaysia
25200 Kuantan, Pahang, MALAYSIA
nurulnajihah@iium.edu.my

Ibrahim Mohamed
Institute of Mathematical Sciences, Faculty of Science, University of Malaya
50603, Kuala Lumpur, MALAYSIA
imohamed@um.edu.my

Ng Kok Haur
Institute of Mathematical Sciences, Faculty of Science, University of Malaya
50603, Kuala Lumpur, MALAYSIA
kokhaur@um.edu.my

Abstract

Recently, there has been a growing interest in integer-valued time series models. In this paper, using a martingale difference, we prove a general theorem on the moment properties of a class of integer-valued time series models. This theorem not only contains results in the recent literature as special cases but also has the advantage of a simpler proof. In addition, we derive the closed form expressions for the kurtosis and skewness of the models. The results are very useful in understanding the behaviour of the processes involved and in estimating the parameters of the models using quadratic estimating functions (QEF). Specifically, we derive the optimal function for the integer-valued GARCH (p, q) known as INGARCH (p, q) model. Simulation study is carried out to compare the performance of QEF estimates with corresponding maximum likelihood (ML) and least squares (LS) estimates for the INGARCH (1,1) model with different sets of parameters. Results show that the QEF estimates produce smaller standard errors than the ML and LS estimates for small sample size and are comparable to the ML estimates for larger sample size. For illustration, we fit the 108 monthly strike data to INGARCH (1, 1) models via QEF, ML and LS methods, and show the applicability of QEF method in practice.

Keyword: Skewness, kurtosis, martingale difference, quadratic estimating functions, integer-valued.

1. Introduction

An increasing number of studies that involve integer-valued time series data can be found in the literature. Zeger (1988) extensively studied the monthly cases of Polio infection in the U.S. from 1970 to 1983. Johansson (1996) considered the effect of lowering speed limits on the number of accidents while Li et al. (2014) investigated the implication of crime cases over time. As a result, there is a need for integer-valued time series models extended to include autoregressive moving average models, the first of which were introduced by Brockwell and Davis (1991) and Emad and Nadjib (1994).

Later, Ferland et al. (2006) extended the classical generalized autoregressive conditional heteroskedastic model with Poisson deviate. To account for overdispersion, Zhu (2011) introduced a new version of Ferland's model with negative binomial deviate. For cases of data with excess zeroes, Zhu (2012) proposed the zero-inflated models with both Poisson and negative binomial deviates. Here, we re-examine some of these models and present simpler derivations of their moment properties using martingale difference. Such martingale difference have been successfully applied to various time series processes, see for example, Thavaneswaran and Abraham (1988) and Ghahramani and Thavaneswaran (2009). These results are very significant for the development of simpler theories on integer-valued time series models, in particular, for estimating the paramaters of the models using the estimating functions method.

The paper is divided as follows: In Section 2, we propose a general class of integer-valued time series models including important models given in Ferland et al. (2006), Zhu (2011) and Zhu (2012). We also derive the basic properties of the model, namely, formulae for the mean, variance, autocovariance and autocorrelation using a new approach, i.e by employing martingale differences. In Section 3, we present the higher order moment properties of the model up to order 4 by using martingale difference. In Section 4, we derive the optimal function for INGARCH (p,q) model via quadratic estimating functions (QEF). Simulation study is conducted to compare the performance of QEF, ML and LS estimates for INGARCH (1,1) model. We illustrate the QEF method in practice to the monthly strike data set given in Jung et al. (2005). Concluding remarks are given in Section 5.

2. Moments of Integer-Valued Time Series Models

Following the notation in Ferland et al. (2006), we consider four types of integer-valued time series model: Poisson (INGARCH), negative binomial (NBINGARCH), zero-inflated Poisson (ZIPINGARCH) and zero-inflated negative binomial (ZINBINGARCH) with conditional mean $E(X_t \mid \Im_{t-1})$ is of the form:

$$E(X_t \mid \Im_{t-1}) = a\lambda_{t,\mathrm{TP}} \quad , \tag{1}$$

$$\lambda_{t,\mathrm{TP}} = \gamma + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j,\mathrm{TP}} \ , \tag{2}$$

where $\Im_{t-1}$ is the σ−field generated by $X_{t-1}, X_{t-2},..., X_1$ with $\lambda_{t,\mathrm{TP}}$ is the intensity parameters with TP = Poisson (P), negative binomial (NB), zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) for the respective models, $\gamma > 0$, $\alpha_i \geq 0$, $i = 1,2,..., p$, and $\beta_j \geq 0$, $j = 1,2,..., q$ and a is the coefficient of the conditional mean with

$$a = \begin{cases} 1 & \text{for INGARCH}(p,q), \\ r & \text{for NBINGARCH}(r,p,q) \text{ by assuming } \lambda_{\mathrm{t.NB}} = \dfrac{1-p_t}{p_t} \\ 1-\omega & \text{for ZIPINGARCH } (\omega, p, q) \text{ and for ZINBGARCH } (\omega, p, q) \end{cases} ,$$

where r is the number of successful trials, p_t is the probability of successful trials and ω is the inflation parameter.

We now apply the martingale difference, $u_t = X_t - E(X_t \mid \Im_{t-1}) = X_t - a\lambda_{t,\mathrm{TP}}$ with $E(u_t) = 0$ and $\mathrm{Var}(u_t) = \sigma_u^2$. Multiplying equation (2) by *a* gives

$$a\lambda_{t,\mathrm{TP}} = a\gamma + a\sum_{i=1}^{p}\alpha_i X_{t-i} + a\sum_{j=1}^{q}\beta_j\lambda_{t-j,\mathrm{TP}} .$$

Since u_t is a martingale difference sequence, the equation can be rewritten in using backward operator, *B*, as

$$\left(1 - a\sum_{i=1}^{p}\alpha_i B^i - \sum_{j=1}^{q}\beta_j B^j\right)X_t = a\gamma + \left(1 - \sum_{j=1}^{q}\beta_j B^j\right)u_t . \quad (3)$$

Now, let $\phi(B) = 1 - a\sum_{i=1}^{p}\alpha_i B^i - \sum_{j=1}^{q}\beta_j B^j$ and $\theta(B) = 1 - \sum_{j=1}^{q}\beta_j B^j$. Then, equation (3) can be represented in the form

$$\phi(B)X_t = a\gamma + \theta(B)u_t . \quad (4)$$

If all the roots of $\phi(B) = 0$ lie outside the unit circle, then the process $\{X_t\}$ is stationary. By letting $\psi(B) = \frac{\theta(B)}{\phi(B)}$ and $\mu = \frac{a\gamma}{\phi(B)}$, the equation (4) can be written as $X_t - \mu = \psi(B)u_t$, i.e.

$$X_t - \mu = \sum_{j=0}^{\infty}\psi_j u_{t-j} , \quad (5)$$

and the variance of X_t is given by

$$\sigma_{X_t}^2 = \sigma_u^2 \sum_{j=0}^{\infty}\psi_j^2 .$$

Using first order stationarity, $E(X_t) = \mu$ and since for large *t*, $E(\lambda_t)$ approaches λ, a constant, then, σ_u^2 can be easily expressed in terms of μ. For the processes INGARCH, NBINGARCH and ZIPINGARCH, it can be easily shown that the corresponding values of σ_u^2 are μ, $\mu\left(1+\frac{\mu}{r}\right)$ and $\mu\left(1+\frac{\omega\mu}{1-\omega}\right)$ respectively. Moreover, for ZINBINGARCH model, σ_u^2 is given by

$$\sigma_u^2 = \begin{cases} \mu\left(1 + a + \frac{\omega\mu}{1-\omega}\right) & \text{for c} = 0 \\ \mu\left(1 + (\omega + a)\frac{\mu}{1-\omega}\right) & \text{for c} = 1 \end{cases},$$

following the index c = 0, 1 appearing in the probability mass function of the zero-inflated negative binomial distribution (see Zhu, 2012). However, we note that the strict stationary properties have been studied only for the INGARCH (*p, q*) model by Ferland et al. (2006). As highlighted by Zhu (2011, 2012), different approaches are required to exhibit the properties for the other three models and are of interest in future work.

The first aim here is to derive the general formula for the first two moments, the autocovariance and the autocorrelation of the integer-valued process $\{X_t\}$ of the form in equations (1-2). The result is given in Theorem 1.

Theorem 1: Under the first and second order stationarity assumptions,

(a) $$E(X_t) = \mu = \frac{a\gamma}{1 - a\sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j},$$

(b) $$\mathrm{Var}(X_t) = \sigma_X^2 = \sigma_u^2 \sum_{j=0}^{\infty}\psi_j^2,$$

(c) $$\gamma_k^X = \sigma_u^2 \sum_{j=0}^{\infty}\psi_j\psi_{j+k},$$

(d) $$\rho_k^X = \frac{\sum_{j=0}^{\infty}\psi_j\psi_{j+k}}{\sum_{j=0}^{\infty}\psi_j^2}.$$

Proof: The mean of X_t can be obtained by taking the expectation of equation (3). Since $E(u_t) = 0$, 1(a) follows. From equation (5), we notice that the process can be represented as a general form of a time series process (see Abraham and Ledolter, 2009), therefore, the variance, autocovariance and correlation of X_t are obtained.

3. Skewness and Kurtosis

In the literature, only the first two moments and the autocovariance are given for integer-valued time series models. In this section, following Thavaneswaran et al. (2005), we obtained the general expression for the skewness and kurtosis for the INGARCH, NBINGARCH, ZIPINGARCH and ZINBINGARCH models.

Theorem 2: Consider a linear stationary process of the form $X_t - \mu = \sum_{j=0}^{\infty}\psi_j u_{t-j}$ where μ is the mean of the random process and u_t is an uncorrelated noise process with mean zero, variance σ_u^2, skewness $\Gamma^{(u)}$ and kurtosis $K^{(u)}$. Define $S_t = (X_t - \mu)^2$. Then, under suitable stationarity conditions, such process will have variance, skewness, kurtosis and correlation given by

(a) $$\mathrm{Var}(S_t) = \left(K^{(u)} - 3\right)\sigma_u^4 \sum_{j=0}^{\infty}\psi_j^4 + 2\sigma_u^4\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2,$$

(b) $$\Gamma^{(X)}=\frac{\sum_{j=0}^{\infty}\psi_j^3\Gamma^{(u)}}{\left(\sum_{j=0}^{\infty}\psi_j^2\right)^{3/2}},$$

(c) $$K^{(X)}=3+\frac{\left(K^{(u)}-3\right)\sum_{j=0}^{\infty}\psi_j^4}{\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2},$$

(d) $$\rho_k^S=\frac{\left(K^{(u)}-3\right)\sum_{j=0}^{\infty}\psi_j^2\psi_{j+k}^2+2\left(\sum_{j=0}^{\infty}\psi_j\psi_{j+k}\right)^2}{\left(K^{(u)}-3\right)\sum_{j=0}^{\infty}\psi_j^4+2\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2}.$$

respectively. The proof is given in Appendix 1.

Example: Using Theorems 1 and 2, we derive the following results for four distributions with p = 1 and q = 1. From equations (1) and (2), the process $\{X_t\}$ is such that

$$E\left(X_t\mid\Im_{t-1}\right)=a\lambda_{t,\mathrm{TP}},$$

$$\lambda_{t,\mathrm{TP}}=\gamma+\alpha_1X_{t-1}+\beta_1\lambda_{t-1,\mathrm{TP}}.$$

It can be shown that the weight ψ_j is given by $\psi_j=a\alpha_1(a\alpha_1+\beta_1)^{j-1}$ where

$$a=\begin{cases}1 & \text{for INGARCH(1,1)}\\ r & \text{for NBINGARCH(1,1) by assuming } \lambda_{\mathrm{t.NB}}=\dfrac{1-p_t}{p_t}\\ 1-\omega & \text{for ZIPINGARCH (1,1) and for ZINBINGARCH (1,1)}\end{cases}.$$

Therefore, the summations of the weights ψ_j are given in the following form:

$$\sum_{j=0}^{\infty}\psi_j^2=\frac{1-2a\alpha_1\beta_1-\beta_1^2}{1-\left(a\alpha_1+\beta_1\right)^2},$$

$$\sum_{j=0}^{\infty}\psi_j^3=\frac{1-3a^2\alpha_1^2\beta_1-3a\alpha_1\beta_1^2-\beta_1^3}{1-\left(a\alpha_1+\beta_1\right)^3},$$

$$\sum_{j=0}^{\infty}\psi_j^4=\frac{1-4a^3\alpha_1^3\beta_1-6a^2\alpha_1^2\beta_1^2-4a\alpha_1\beta_1^3-\beta_1^4}{1-\left(a\alpha_1+\beta_1\right)^4},$$

$$\sum_{j=0}^{\infty} \psi_j \psi_{j+k} = a\alpha_1 (a\alpha_1 + \beta_1)^{k-1} \left\{ \frac{1 - a\alpha_1\beta_1 - \beta_1^2}{1 - (a\alpha_1 + \beta_1)^2} \right\},$$

and

$$\sum_{j=0}^{\infty} \psi_j^2 \psi_{j+k}^2 = a^2\alpha_1^2 (a\alpha_1 + \beta_1)^{2k-2} \left\{ \frac{1 - (a\alpha_1 + \beta_1)^4 + a^2\alpha_1^2 (a\alpha_1 + \beta_1)^2}{1 - (a\alpha_1 + \beta_1)^4} \right\}.$$

From Theorem 1(c), the autocovariance of the $\{X_t\}$ process with order (1,1) can be written as

$$\gamma_k^X = \sigma_u^2 a\alpha_1 (a\alpha_1 + \beta_1)^{k-1} \left\{ \frac{1 - (a\alpha_1 + \beta_1)\beta_1}{1 - (a\alpha_1 + \beta_1)^2} \right\},$$

and from Theorem 1(b), the variance of the $\{X_t\}$ process is

$$\sigma_X^2 = \sigma_u^2 \left\{ \frac{1 - (a\alpha_1 + \beta_1)^2 + a^2\alpha_1^2}{1 - (a\alpha_1 + \beta_1)^2} \right\},$$

while from Theorem 1(d), the correlation of the $\{X_t\}$ process is

$$\rho_k^X = \frac{a\alpha_1 (a\alpha_1 + \beta_1)^{k-1} [1 - (a\alpha_1 + \beta_1)\beta_1]}{1 - (a\alpha_1 + \beta_1)^2 + (a\alpha_1)^2},$$

where a is as defined earlier for the different models.

Using the same arguments as in Section 2, we can find the skewness and kurtosis of u_t. The skewness $\Gamma^{(u)}$ for INGARCH, NBINGARCH and ZIPINGARCH is $\frac{1}{\sqrt{\mu}}$, $\frac{r + 2\mu}{\sqrt{r\mu(r+\mu)}}$ and $\frac{\omega(1+2\omega)\mu^2 + 3\omega\mu(1-\omega) + (1-\omega)^2}{\sqrt{\mu(1-\omega)}(1-\omega+\omega\mu)^{3/2}}$ respectively. On the other hand, the kurtosis, $K^{(u)}$ for INGARCH, NBINGARCH and ZIPINGARCH is respectively $\frac{1}{\mu}$, $\frac{r^2 - 2\mu^2 - 2r\mu}{r\mu(r+\mu)}$ and K where K is

$$\frac{(1-\omega)^3 + \omega\mu\left(7 - 14\omega + 7\omega^2 + 6\mu + \mu^2 + 18\mu\omega - 12\omega^2\mu - 6\mu^2\omega + 6\mu^2\omega^2\right)}{(1-\omega)\mu(1+\omega\mu)^2}.$$

However, the similar corresponding expression for ZINBINGARCH is complicated but can still be solved using standard mathematical software.

4. General Theory of Quadratic Estimating Functions

Godambe (1960) was the first to introduce regular estimating functions (EF) that satisfy certain conditions and procedures for choosing an optimal EF. The requirement for a regular EF, $g(X_t;\boldsymbol{\theta})$ are:

(i) $E[g(X_t;\boldsymbol{\theta})] = \int g(X_t;\boldsymbol{\theta}) f(X_t;\boldsymbol{\theta}) dX_t = 0$,

(ii) $\dfrac{\partial g(X_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ exists for all $\boldsymbol{\theta} \in \Theta$ where Θ is the parameter space,

(iii) $\int g(X_t;\boldsymbol{\theta}) f(X_t;\boldsymbol{\theta}) dX_t$ is differentiable under the sign of integration,

(iv) $E\left[\dfrac{\partial g(X_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^2 > 0$ for all $\boldsymbol{\theta} \in \Theta$,

(v) $\mathrm{Var}[g(X_t;\boldsymbol{\theta})] = E\left[g^2(X_t;\boldsymbol{\theta})\right] < \infty$.

According to Godambe (1960), to find the optimal EF, say $g^*(X_t;\boldsymbol{\theta})$ two criteria should be satisfied. First, the estimated parameter should be as close as possible to the true value. This means that the variance $\mathrm{Var}[g(X_t;\boldsymbol{\theta})] = E\left[g^2(X_t;\boldsymbol{\theta})\right]$ should be minimized and therefore $E\left[g^{*2}(X_t;\boldsymbol{\theta})\right] \le E\left[g^2(X_t;\boldsymbol{\theta})\right]$. The second criterion is that the expected values of the derivatives of the function $g(X_t;\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, $\left\{E\left[\dfrac{\partial g(X_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\right\}$ should be large as possible i.e. $\left\{E\left[\dfrac{\partial g^*(X_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\right\} \ge \left\{E\left[\dfrac{\partial g(X_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\right\}$. By following both criteria, the optimal EF, $g^*(X_t;\boldsymbol{\theta})$ can be defined as follows:

Definition 1

Let G denote the class of all regular EFs. The $g^*(X_t;\boldsymbol{\theta}) \in G$ is said to be optimal if

$$\frac{E\left(g^{*2}(X_t;\boldsymbol{\theta})\right)}{\left\{E\left[\dfrac{\partial g^*(X_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\right\}^2} \le \frac{E\left[g^2(X_t;\boldsymbol{\theta})\right]}{\left\{E\left[\dfrac{\partial g(X_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\right\}^2}.$$

Further, Godambe (1985) studied the inference of discrete time series processes using estimating functions. He considered a class $\left\{g : g(\boldsymbol{\theta}) = \sum_{t=1}^{n} a_{t-1} h_t\right\}$ of EF which is linear combination of h_t's where the expected value $E(h_t \mid \Im_{t-1}) = 0$ since $\Im_{t-1}$ is the $\sigma-$field generated by $\{X_s; s \le t-1\}$. The theorem below is the result in Godambe (1985) on optimal EFs for the dependent case.

Theorem 3: Let $\left\{g : g(\boldsymbol{\theta}) = \sum_{t=1}^{n} a_{t-1} h_t\right\}$ be the class of all EFs where h_t and a_{t-1} are assumed to be differentiable with respect to $\boldsymbol{\theta}$ for $t = 1,2,...,n$. Then, the optimal estimating function $g^*(\boldsymbol{\theta})$ minimizing $\dfrac{E\left(g^2(X_t;\boldsymbol{\theta})\right)}{\left\{E\left[\dfrac{\partial g(X_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\right\}^2}$ is $g^*(\boldsymbol{\theta}) = \sum_{t=1}^{n} a_{t-1}^* h_t$ where

$$a_{t-1}^* = E\left[\frac{\partial h_t}{\partial \boldsymbol{\theta}} \middle| \Im_{t-1}\right] \Big/ E\left[h_t^2 \middle| \Im_{t-1}\right].$$

The EF method was later extended by Liang et al. (2011) to the case where the first four conditional moments are known. The functions used are called quadratic estimating functions (QEF).

Now, we assume that the discrete time stochastic process $\{X_t, t = 1,2,...,n\}$ has the following conditional moments depending only on the parameter $\boldsymbol{\theta}$:

$$\mu_t(\boldsymbol{\theta}) = \mu_t = E\left[X_t \middle| \Im_{t-1}\right],$$

$$\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2 = E\left[(X_t - \mu_t(\boldsymbol{\theta}))^2 \middle| \Im_{t-1}\right],$$

$$\Gamma_t(\boldsymbol{\theta}) = \Gamma_t = \frac{1}{\sigma_t^3(\boldsymbol{\theta})} E\left[(X_t - \mu_t(\boldsymbol{\theta}))^3 \middle| \Im_{t-1}\right],$$

$$\kappa_t(\boldsymbol{\theta}) = \kappa_t = \frac{1}{\sigma_t^4(\boldsymbol{\theta})} E\left[(X_t - \mu_t(\boldsymbol{\theta}))^4 \middle| \Im_{t-1}\right] - 3.$$

We estimate the parameter $\boldsymbol{\theta}$ using two classes of martingale differences

$$\{m_t(\boldsymbol{\theta}) = m_t = X_t - \mu_t, t = 1,2,...,n\}, \quad (6)$$

and $$\{s_t(\boldsymbol{\theta}) = s_t = m_t^2 - \sigma_t^2, t = 1,2,...,n\}, \quad (7)$$

such that

$$\langle m \rangle_t = E\left[m_t^2 \middle| \Im_{t-1}\right] = \sigma_t^2, \quad (8)$$

$$\langle s \rangle_t = E\left[s_t^2 \middle| \Im_{t-1}\right] = \sigma_t^4(\kappa_t + 2), \quad (9)$$

$$\langle m, s \rangle_t = E\left[m_t s_t \middle| \Im_{t-1}\right] = \sigma_t^3 \Gamma_t. \quad (10)$$

The following theorem is obtained from Liang et al. (2011).

Theorem 4: In the class $G_Q = \left\{ g_Q(\boldsymbol{\theta}) : g_Q(\boldsymbol{\theta}) = \sum_{t=1}^{n} (a_{t-1} m_t + b_{t-1} s_t) \right\}$ of all quadratic estimating functions, the optimal estimating functions is given by $g_Q^*(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(a_{t-1}^* m_t + b_{t-1}^* s_t \right)$ where

$$a_{t-1}^* = \left(1 - \frac{\langle m,s \rangle_t^2}{\langle m \rangle_t \langle s \rangle_t} \right)^{-1} \left(-\frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{\langle m,s \rangle_t}{\langle m \rangle_t \langle s \rangle_t} \right),$$

and $$b_{t-1}^* = \left(1 - \frac{\langle m,s \rangle_t^2}{\langle m \rangle_t \langle s \rangle_t} \right)^{-1} \left(\frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{\langle m,s \rangle_t}{\langle m \rangle_t \langle s \rangle_t} - \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{1}{\langle s \rangle_t} \right).$$

4.1 The INGARCH (*p*, *q*) Model

In this subsection, we focus on INGARCH (*p*, *q*) model given by

$$E(X_t \mid \Im_{t-1}) = \lambda_{t,\mathrm{P}},$$

$$\lambda_{t,\mathrm{P}} = \gamma + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j,\mathrm{P}}.$$

The parameters of interest are $\boldsymbol{\theta} = \left(\gamma_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \right)'$. The mean, variance, skewness and kurtosis of X_t conditional on $\Im_{t-1}$ are

$$\mu_t = \lambda_{t,\mathrm{P}} = \lambda_{t,\mathrm{P}}(\boldsymbol{\theta}), \quad \sigma_t^2 = \lambda_{t,\mathrm{P}}, \quad \Gamma_t = \frac{1}{\sqrt{\lambda_{t,\mathrm{P}}}}, \quad \text{and } \kappa_t = \frac{1}{\lambda_{t,\mathrm{P}}}.$$

From equations (6) and (7), the martingale differences considered are $m_t = X_t - \lambda_{t,\mathrm{P}}$ and $s_t = (X_t - \lambda_{t,\mathrm{P}})^2 - \lambda_{t,\mathrm{P}}$. On the other hand, the conditional expectations in equations (8), (9) and (10) are $\langle m \rangle_t = \lambda_{t,\mathrm{P}}$, $\langle s \rangle_t = \lambda_{t,\mathrm{P}}(1 + 2\lambda_{t,\mathrm{P}})$ and $\langle m,s \rangle_t = \lambda_{t,\mathrm{P}}$. Let $C_t = \left(1 - \frac{\langle m,s \rangle_t^2}{\langle m \rangle_t \langle s \rangle_t} \right)^{-1} = \frac{1 + 2\lambda_{t,\mathrm{P}}}{2\lambda_{t,\mathrm{P}}}$. Then, from Theorem 4, the optimal QEF is $g_Q^*(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(a_{t-1}^* m_t + b_{t-1}^* s_t \right)$ where

$$\begin{aligned} a_{t-1}^* &= C_t \left(-\frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{\langle m,s \rangle_t}{\langle m \rangle_t \langle s \rangle_t} \right) \\ &= \frac{1 + 2\lambda_{t,\mathrm{P}}}{2\lambda_{t,\mathrm{P}}} \left[-\lambda'_{t,\mathrm{P}} \frac{1}{\lambda_{t,\mathrm{P}}} + \lambda'_{t,P} \frac{\lambda_{t,\mathrm{P}}}{(1 + 2\lambda_{t,\mathrm{P}})[\lambda_{t,\mathrm{P}}]^2} \right] \\ &= -\frac{\lambda'_{t,\mathrm{P}}}{\lambda_{t,\mathrm{P}}}, \end{aligned}$$

and

$$
\begin{aligned}
b_{t-1}^{*} &= C_t\left(\frac{\partial \mu_t}{\partial \boldsymbol{\theta}}\frac{\langle m,s\rangle_t}{\langle m\rangle_t\langle s\rangle_t} - \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}\frac{1}{\langle s\rangle_t}\right)\\
&= \frac{1+2\lambda_{t,\mathrm{P}}}{2\lambda_{t,\mathrm{P}}}\left[\lambda'_{t,\mathrm{P}}\frac{1}{\lambda_{t,\mathrm{P}}\left(1+2\lambda_{t,\mathrm{P}}\right)} - \lambda'_{t,\mathrm{P}}\left(\frac{1}{\lambda_{t,\mathrm{P}}\left(1+2\lambda_{t,\mathrm{P}}\right)}\right)\right]\\
&= 0.
\end{aligned}
$$

Therefore, one can conclude that, if the conditional mean and conditional variance are the same, the QEF method can be reduced to the EF method. Since we have $p+q+1$ parameters, then $\lambda'_{t,\mathrm{P}}(\boldsymbol{\theta})$ are

$$
\begin{aligned}
\lambda'_{t,\mathrm{P}}(\boldsymbol{\theta}) &= \left(\frac{\partial \lambda_{t,\mathrm{P}}}{\partial \gamma}, \frac{\partial \lambda_{t,\mathrm{P}}}{\partial \alpha_i}, \frac{\partial \lambda_{t,\mathrm{P}}}{\partial \beta_j}\right)'\\
&= \left(1+\sum_{j=1}^{q}\beta_j\frac{\partial \lambda_{t-j,\mathrm{P}}}{\partial \gamma},\ X_{t-i}+\sum_{j=1}^{q}\beta_j\frac{\partial \lambda_{t-j,\mathrm{P}}}{\partial \alpha_i}, \lambda_{t-j,P}+\sum_{k=1}^{q}\beta_k\frac{\partial \lambda_{t-k,\mathrm{P}}}{\partial \beta_j}\right)',
\end{aligned}
$$

where $i=1,2,\ldots,p$ and $j=1,2,\ldots,q$. Hence, the optimal QEF for each parameter are:

$$
g_Q^{*}(\gamma) = \sum_{t=1}^{n}\frac{1}{\lambda_{t,\mathrm{P}}}\left(1+\sum_{j=1}^{q}\beta_j\frac{\partial \lambda_{t-j,\mathrm{P}}}{\partial \gamma}\right)\left(X_t-\lambda_{t,\mathrm{P}}\right), \tag{11}
$$

$$
g_Q^{*}(\alpha_i) = \sum_{t=1}^{n}\frac{1}{\lambda_{t,\mathrm{P}}}\left(X_{t-i}+\sum_{j=1}^{q}\beta_j\frac{\partial \lambda_{t-j,\mathrm{P}}}{\partial \alpha_i}\right)\left(X_t-\lambda_{t,\mathrm{P}}\right), \tag{12}
$$

$$
g_Q^{*}(\beta_j) = \sum_{t=1}^{n}\frac{1}{\lambda_{t,\mathrm{P}}}\left(\lambda_{t-j,\mathrm{P}}+\sum_{k=1}^{q}\beta_k\frac{\partial \lambda_{t-k,\mathrm{P}}}{\partial \beta_j}\right)\left(X_t-\lambda_{t,\mathrm{P}}\right). \tag{13}
$$

The formulation of optimal equations for the other three models which are NBINGARCH (*p*,*q*), ZIPINGARCH (*p*,*q*) and ZINBINGARCH (p,q) models are the same as INGARCH (*p*,*q*) model. The optimal equation(s) for the additional extra parameter (s) in the above three models can be derived using optimal estimating functions in Theorem 4. An optimal estimate of $\boldsymbol{\theta}$ can be obtained by solving the equation(s) $g_Q^{*}(\boldsymbol{\theta})=0$.

4.2 Simulation Study

Let *N* and *n* be the number of simulations and the sample size generated respectively from the INGARCH (1, 1) models given by

Model 1: $E\left(X_t \mid \Im_{t-1}\right)=\lambda_{t,\mathrm{P}}$; $\lambda_{t,\mathrm{P}} = 0.2+0.4X_{t-1}+0.1\lambda_{t-1,\mathrm{P}}$

Model 2: $E\left(X_t \mid \Im_{t-1}\right)=\lambda_{t,\mathrm{P}}$; $\lambda_{t,\mathrm{P}} = 0.1+0.6X_{t-1}+0.3\lambda_{t-1,\mathrm{P}}$

Model 3: $E\left(X_t \mid \Im_{t-1}\right)=\lambda_{t,\mathrm{P}}$; $\lambda_{t,\mathrm{P}} = 0.3+0.4X_{t-1}+0.2\lambda_{t-1,\mathrm{P}}$

Here, we demonstrate how to estimate the parameters using the QEF method:

- *Step 1-* *Generate the data*: We first generate the data from given true values. Then, we choose the observations numbering from 100 to 100 + *n*.
- *Step 2-* *Initialize the parameters*: We set the initial values for α_1 and β_1 by taking the typical values for each parameter, namely $\alpha_1 = 0.1$ and $\beta_1 = 0.8$. On the other hand, we take the value of γ to be the mean μ_X, of the generated data in Step 1, namely, $\gamma = 0.1\mu_X$, (see Ferland et al., 2006).
- *Step 3-* *Estimate the parameters*: Using *nleqslv*, we solve the simultaneous optimal equations (11) to (13) in R-cran programming language in order to obtain the QEF estimates of γ, α_1 and β_1 for the INGARCH (1,1) model.

Table 1: Simulation results for INGARCH (1,1) with $\gamma = 0.2$, $\alpha_1 = 0.4$ and $\beta_1 = 0.1$.

		n = 100			n = 250			n = 500		
		ML	LS	QEF	ML	LS	QEF	ML	LS	QEF
$\hat{\gamma}$	Mean	0.167	0.241	0.170	0.198	0.185	0.185	0.202	0.190	0.191
	Bias	-0.032	-0.159	-0.030	-0.002	-0.015	-0.015	0.002	-0.010	-0.010
	SE	0.067	0.101	0.057	0.049	0.048	0.046	0.035	0.037	0.034
	RMSE	0.074	0.108	0.065	0.049	0.048	0.046	0.035	0.058	0.035
$\hat{\alpha}_1$	Mean	0.346	0.367	0.339	0.387	0.366	0.375	0.393	0.379	0.386
	Bias	-0.054	-0.033	-0.062	-0.012	-0.034	-0.025	-0.007	-0.021	-0.014
	SE	0.126	0.145	0.129	0.089	0.098	0.089	0.060	0.062	0.060
	RMSE	0.136	0.149	0.142	0.090	0.098	0.092	0.061	0.065	0.061
$\hat{\beta}_1$	Mean	0.228	0.004	0.216	0.131	0.162	0.155	0.096	0.144	0.132
	Bias	0.128	-0.096	0.116	0.013	0.062	0.055	-0.004	0.044	0.032
	SE	0.185	0.265	0.145	0.131	0.116	0.111	0.092	0.092	0.086
	RMSE	0.225	0.282	0.180	0.132	0.132	0.124	0.092	0.102	0.092

Discussion

In order to evaluate the performance of the QEF, maximum likelihood (ML) and least squares (LS) methods for INGARCH (1,1) model, a simulation study was carried out with $N = 500$, $n = 100, 250, 500, 1000, 1500$. The performance of each parameter estimates is measured using bias, standard error (SE) and root mean squared error (RMSE). The results are shown in Tables 1 to 3.

Table 1: Simulation results for INGARCH (1,1) with $\gamma = 0.2$, $\alpha_1 = 0.4$ and $\beta_1 = 0.1$ (Cont.).

		n = 1000			*n* = 1500		
		ML	LS	QEF	ML	LS	QEF
$\hat{\gamma}$	Mean	0.201	0.196	0.197	0.201	0.197	0.198
	Bias	0.001	-0.005	-0.003	0.001	-0.003	-0.002
	SE	0.026	0.028	0.026	0.023	0.025	0.022
	RMSE	0.027	0.028	0.027	0.023	0.025	0.022
$\hat{\alpha}_1$	Mean	0.399	0.395	0.396	0.399	0.396	0.399
	Bias	-0.001	-0.005	-0.004	-0.001	-0.004	-0.001
	SE	0.041	0.045	0.042	0.031	0.037	0.030
	RMSE	0.041	0.045	0.043	0.033	0.038	0.031
$\hat{\beta}_1$	Mean	0.097	0.116	0.112	0.097	0.109	0.102
	Bias	-0.003	0.016	0.012	-0.003	0.009	0.002
	SE	0.072	0.073	0.066	0.054	0.064	0.049
	RMSE	0.072	0.075	0.067	0.055	0.064	0.050

Table 2: Simulation results for INGARCH (1,1) with $\gamma = 0.1$, $\alpha_1 = 0.6$ and $\beta_1 = 0.3$.

		n = 100			*n* = 250			*n* = 500		
		ML	LS	QEF	ML	LS	QEF	ML	LS	QEF
$\hat{\gamma}$	Mean	0.137	0.169	0.099	0.117	0.131	0.111	0.109	0.118	0.104
	Bias	0.037	0.069	-0.001	0.017	0.031	0.011	0.009	0.018	0.004
	SE	0.085	0.100	0.055	0.041	0.052	0.037	0.026	0.034	0.023
	RMSE	0.093	0.122	0.055	0.044	0.061	0.038	0.027	0.039	0.024
$\hat{\alpha}_1$	Mean	0.564	0.533	0.542	0.591	0.569	0.584	0.595	0.583	0.591
	Bias	-0.036	-0.067	-0.058	-0.009	-0.031	-0.016	-0.005	-0.017	-0.009
	SE	0.128	0.162	0.105	0.089	0.101	0.088	0.059	0.073	0.058
	RMSE	0.132	0.175	0.189	0.089	0.105	0.089	0.059	0.075	0.059
$\hat{\beta}_1$	Mean	0.272	0.227	0.282	0.277	0.282	0.287	0.286	0.288	0.294
	Bias	-0.027	-0.073	-0.018	-0.023	-0.018	-0.013	-0.013	-0.012	-0.006
	SE	0.129	0.218	0.127	0.094	0.114	0.093	0.062	0.082	0.062
	RMSE	0.132	0.230	0.149	0.097	0.115	0.094	0.063	0.083	0.062

Table 2: Simulation results for INGARCH (1,1) with $\gamma = 0.1$, $\alpha_1 = 0.6$ and $\beta_1 = 0.3$ (Cont.).

		n = 1000			*n* = 1500		
		ML	LS	QEF	ML	LS	QEF
$\hat{\gamma}$	Mean	0.104	0.111	0.102	0.103	0.108	0.102
	Bias	0.004	0.011	0.002	0.003	0.008	0.002
	SE	0.016	0.024	0.016	0.013	-0.020	0.009
	RMSE	0.017	0.026	0.017	0.014	0.022	0.011
$\hat{\alpha}_1$	Mean	0.598	0.594	0.596	0.595	0.597	0.598
	Bias	-0.002	-0.01	-0.004	-0.005	-0.003	-0.002
	SE	0.038	0.051	0.040	0.033	0.044	0.025
	RMSE	0.038	0.051	0.040	0.033	0.044	0.026
$\hat{\beta}_1$	Mean	0.294	0.289	0.297	0.294	0.291	0.297
	Bias	-0.006	-0.01	-0.003	-0.005	-0.009	-0.003
	SE	0.041	0.056	0.043	0.034	0.048	0.028
	RMSE	0.042	0.057	0.043	0.034	0.049	0.030

Table 3: Simulation results for INGARCH (1,1) with $\gamma = 0.3$, $\alpha_1 = 0.4$ and $\beta_1 = 0.2$.

		n = 100			*n* = 250			*n* = 500		
		ML	LS	QEF	ML	LS	QEF	ML	LS	QEF
$\hat{\gamma}$	Mean	0.299	0.248	0.275	0.317	0.304	0.299	0.314	0.307	0.306
	Bias	-0.001	-0.052	-0.025	0.017	0.004	-0.001	0.014	0.007	0.006
	SE	0.113	0.115	0.101	0.082	0.081	0.074	0.060	0.063	0.060
	RMSE	0.113	0.126	0.104	0.084	0.081	0.074	0.062	0.063	0.060
$\hat{\alpha}_1$	Mean	0.379	0.366	0.379	0.398	0.382	0.388	0.399	0.392	0.395
	Bias	-0.021	-0.034	-0.021	-0.002	-0.018	-0.012	-0.001	-0.008	-0.005
	SE	0.106	0.145	0.108	0.072	0.075	0.070	0.050	0.055	0.050
	RMSE	0.109	0.149	0.113	0.072	0.081	0.071	0.050	0.055	0.050
$\hat{\beta}_1$	Mean	0.226	0.109	0.236	0.178	0.209	0.210	0.180	0.201	0.194
	Bias	0.026	-0.091	0.036	-0.022	0.009	0.010	-0.020	0.001	-0.006
	SE	0.163	0.269	0.153	0.128	0.125	0.119	0.098	0.098	0.095
	RMSE	0.164	0.284	0.163	0.130	0.125	0.119	0.099	0.099	0.095

Table 3: Simulation results for INGARCH (1,1) with $\gamma = 0.3$, $\alpha_1 = 0.4$ and $\beta_1 = 0.2$ (Cont.).

		n = 1000			n = 1500		
		ML	LS	QEF	ML	LS	QEF
$\hat{\gamma}$	Mean	0.309	0.308	0.307	0.306	0.305	0.304
	Bias	0.009	0.008	0.007	0.006	0.005	0.004
	SE	0.046	0.051	0.044	0.038	0.042	0.034
	RMSE	0.046	0.051	0.045	0.038	0.042	0.035
$\hat{\alpha}_1$	Mean	0.401	0.399	0.400	0.401	0.398	0.401
	Bias	0.001	-0.001	0.000	0.001	-0.002	0.001
	SE	0.034	0.038	0.034	0.028	0.032	0.027
	RMSE	0.034	0.038	0.034	0.028	0.032	0.029
$\hat{\beta}_1$	Mean	0.186	0.185	0.189	0.190	0.192	0.192
	Bias	-0.014	0.015	-0.011	-0.010	-0.008	-0.008
	SE	0.074	0.079	0.070	0.059	0.066	0.052
	RMSE	0.075	0.080	0.071	0.059	0.066	0.055

A number of interesting results can be highlighted. Firstly, for the small sample sizes, $n = 100, 250, 500,$ the QEF estimates give the smaller values of SE and RMSE compared to other two methods. However, as n increases, the SE and RMSE values for the QEF estimates are always marginal smaller than ML and LS estimates. Secondly, as expected, as n increases from 100 to 1500, all the SE and RMSE of QEF, ML and LS estimates are consistently decreases. Lastly, it is important to point out the computational times for the QEF method is four times shorter than the ML method and three times shorter than the LS method when the simulation is done using R-cran programming. The R codes are available upon request.

Therefore, we can conclude that the QEF method provided consistently accurate estimates and computation effective than the ML and LS methods in the parameter estimation of INGARCH models.

4.3 Real data Example

We apply the proposed methodology to analyze the 108 monthly strike data from January 1994 to December 2002 given by Jung et al. (2005). The data are available at the U.S. Bureau of Labor Statistics (http://www.bls.gov/wsp/) (see Weiβ, 2010). It describes the number of work stoppages leading to 1000 workers or more effectively idle during the period. The time series is given in Figure 1.

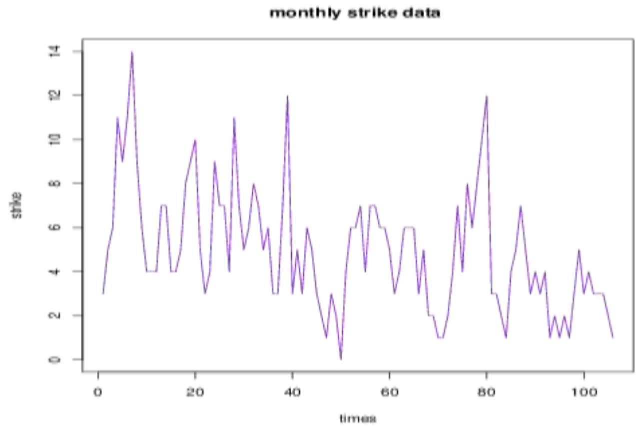

Figure 1: The monthly strike data from January 1994 to December 2002.

We fit the data using the INGARCH (1,1) model via the QEF, ML and LS methods. Then we obtain the parameter estimates together with their respective standard errors in parenthesis are shown in Table 4. We observe that, the QEF and ML methods give the same values of estimates. As expected, the standard errors of the QEF estimates are the smallest as compared to other two methods.

Table 4: The estimated parameters of INGARCH (1,1) model. Values in parentheses are standard errors of parameter estimates.

Method	$\hat{\gamma}$	$\hat{\alpha}_1$	$\hat{\beta}_1$
QEF	1.623 (0.428)	0.610 (0.081)	0.064 (0.114)
ML	1.623 (0.502)	0.610 (0.095)	0.064 (0.125)
LS	1.854 (0.512)	0.596 (0.112)	0.032 (0.128)

Then, to investigate the model fitting adequacy, we consider the Pearson residual defined by $z_t = \dfrac{X_t - \lambda_{t,\mathrm{P}}(\hat{\theta})}{\sqrt{\lambda_{t,\mathrm{P}}(\hat{\theta})}}$. According to Kedeem and Fakianos (2002), for the specified model, the sequence z_t should have mean and variance close to 0 and 1 respectively and the sequence does not have serial correlation. We found that in our case, the mean and variance of the Pearson residuals are 0.032 and 1.009 respectively and are thus close to zero and unity as desired. Moreover, by using Ljung-Box (LB) statistics, the results from Table 5 indicate that there is no significant serial correlation in the residual.

Table 5: Diagnostics for INGARCH (1,1) model using QEF method.

	$LB(z_t)$	$LB(z_t^2)$
χ^2	28.1	21.3
p-value	0.565	0.878

Furthermore, to examine the model adequacy, from Figure 2(a), there is no trend observed indicating the randomness of the residuals and in Figure 2(b), the plot does not exceed the dotted line. Therefore, the INGARCH (1, 1) model is a good fit for the monthly strike data given in Jung et al. (2005).

5. Concluding Remarks

This paper studied the moments of four integer-valued time series models, namely, the Poisson, negative binomial, zero-inflated Poisson and zero-inflated negative binomial models. We used the martingale difference to derive the higher order moments of all four models. The results for the first two moments are similar to those found in Zhu (2011) and Zhu (2012), but the derivation was much simpler. In addition, we derived the higher order moments of integer-valued time series up to order 4. However, the results hold for only the INGARCH (p, q) model. Further investigations on the stationarity of the other three models are required. Furthermore, we developed the quadratic estimating functions method mainly focusing on the INGARCH (p, q) model.

To investigate the performance of the QEF method compared to those of the LS and ML methods, simulation was carried out to obtain the estimated parameters together with their standard errors. The results showed that the QEF estimates give smaller standard errors and computation effective compared to the ML and LS estimates. Lastly, we model the monthly strike data using the INGARCH (1,1) model via QEF method. The adequacy of fit was investigated using diagnostic tools based on the Pearson residuals. For the future research, other estimation methods such as Kalman filter can be considered.

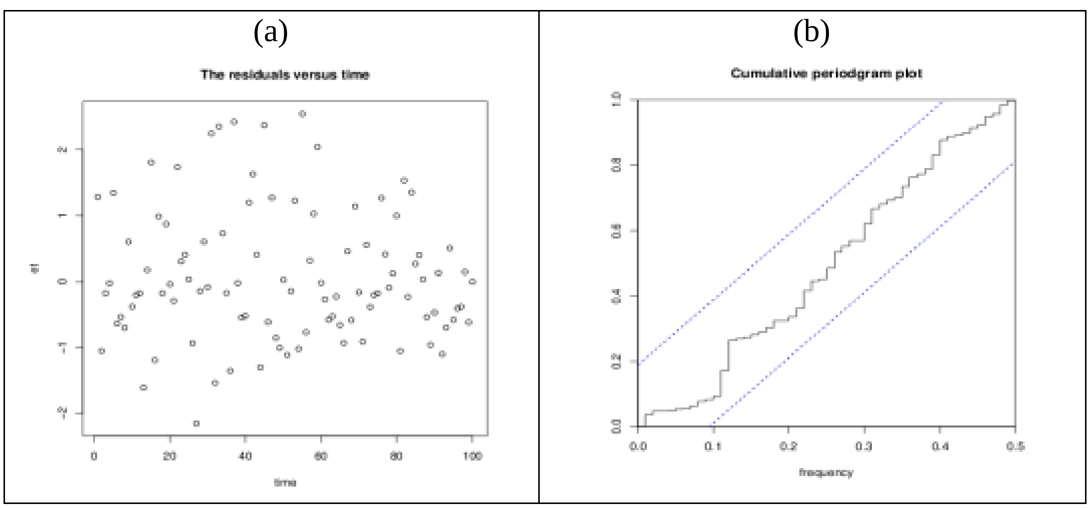

Figure 2: (a) The Pearson residual plot. (b) The periodagram plot.

Acknowledgement

This research is financially supported by Fundamental Research Grant Scheme No. FP012-2013A, University of Malaya Research Grant No. RP014C-15SUS and IIUM Research Initiative Grant Scheme No. RIGS16-311-0475.

References

1. Abraham, B., and Ledolter, J. (2009). Statistical Methods for Forecasting. John Wiley and Sons.
2. Aly, E.E.A.A., and Bouzar, N. (1994). On some integer-valued autoregressive moving average models. Journal of Multivariate Analysis, 50(1), 132-151.
3. Brockwell, P.J., and Davis, R.A. (1991) Time Series: Theory and Methods, 2nd ed. Springer, New York.
4. Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. Journal of Time Series Analysis, 27(6), 923-942.
5. Ghahramani, M. and Thavaneswaran, A. (2009). On some properties of autoregressive conditional Poisson (ACP) models. Economics Letters, 105(3), 273-275.
6. Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. Annals of Mathematical Statistics, 31(4), 1208-1212.
7. Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes. Biometrika, 72(2), 419-428.
8. Johansson, P. (1996). Speed limitation and motorway casualties: A time series count data regression approach. Accident Analysis & Prevention, 28(1), 73-87.
9. Jung, R. C., Ronning, G., and Tremayne, A. R. (2005). Estimation in conditional first order autoregression with discrete support. Statistical Papers, 46(2), 195-224.
10. Kedeem, B., and Fokianos, K. (2002). Regression Models for Time Series Analysis. John Wiley and Sons.
11. Li, G., Haining, R., Richardson, S., and Best, N. (2014). Space-time variability in burglary risk: A Bayesian spatio-temporal modelling approach. Spatial Statistics, 9, 180-191.
12. Liang, Y., Thavaneswaran, A., and Abraham, B. (2011). Joint estimation using quadratic estimating function. Journal of Probability and Statistics. (http://dx.doi.org/10.1155/2011/372512).
13. Thavaneswaran, A., and Abraham, B. (1988). Estimation for non-linear time series models using estimating equations. Journal of Time Series Analysis, 9(1), 99-108.
14. Thavaneswaran, A., Appadoo, S.S. and Peiris, S. (2005). Forecasting volatility. Statistics and Probability Letters, 75(1), 1-10.
15. Weiβ, C. H. (2010). The INARCH(1) model for overdispersed time series of counts. Communications in Statistics - Simulation and Computation, 39(6), 1269-1291.
16. Zeger, S.L. (1988). A regression model for time series of counts. Biometrica, 74(4), 621-629.

17. Zhu, F. (2011). A negative binomial integer-valued GARCH model. Journal of Time Series Analysis, 32(1), 54-67.

18. Zhu, F. (2012). Zero-inflated Poisson and negative binomial integer-valued GARCH models. Journal of Statistical Planning and Inference, 142(4), 826-839.

Appendix 1

(a) Let $S_t = (X_t - \mu)^2$. Then,

$$E(S_t) = E\left[(X_t - \mu)^2\right] = \text{Var}(X_t) = \sigma_u^2 \sum_{j=0}^{\infty} \psi_j^2,$$

and

$$E\left(S_t^2\right) = E\left[(X_t - \mu)^4\right] = E\left[\left(\sum_{j=0}^{\infty} \psi_j u_{t-j}\right)^4\right],$$

$$= \sigma_u^4 \sum_{j=0}^{\infty} \psi_j^4 K^{(u)} + 6\sigma_u^4 \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} \psi_i^2 \psi_j^2,$$

Since from the multinomial expansion, $E(u_t) = 0$ and $E\left(u_t^4\right) = K^{(u)}\sigma_u^4$. Similarly, we can show that

$$6\sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} \psi_i^2 \psi_j^2 = 3\left(\sum_{j=0}^{\infty} \psi_j^2\right)^2 - 3\sum_{j=0}^{\infty} \psi_j^4.$$

Hence,

$$\begin{aligned}
\text{Var}\left[(X_t - \mu)^2\right] &= \text{Var}(S_t) \\
&= E\left(S_t^2\right) - \left[E(S_t)\right]^2 \\
&= \sigma_u^4 \sum_{j=0}^{\infty} \psi_j^4 K^{(u)} + 6\sigma_u^4 \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} \psi_i^2 \psi_j^2 - \left(\sigma_u^2 \sum_{j=0}^{\infty} \psi_j^2\right)^2 \\
&= \sigma_u^4 \sum_{j=0}^{\infty} \psi_j^4 K^{(u)} + 3\sigma_u^4 \left(\sum_{j=0}^{\infty} \psi_j^2\right)^2 - 3\sigma_u^4 \sum_{j=0}^{\infty} \psi_j^4 - \sigma_u^4 \left(\sum_{j=0}^{\infty} \psi_j^2\right)^2 \\
&= \left(K^{(u)} - 3\right)\sigma_u^4 \sum_{j=0}^{\infty} \psi_j^4 + 2\sigma_u^4 \left(\sum_{j=0}^{\infty} \psi_j^2\right)^2.
\end{aligned}$$

(b) Using the third moment and Theorem 1(b), the skewness of X_t is

$$\Gamma^{(X)} = \frac{E\left[(X_t - \mu)^3\right]}{\left[\text{Var}(X_t)\right]^{3/2}}$$

$$= \frac{\sigma_u^3 \sum_{j=0}^{\infty} \psi_j^3 \Gamma^{(u)}}{\left(\sigma_u^2 \sum_{j=0}^{\infty} \psi_j^2\right)^{3/2}} = \frac{\Gamma^{(u)} \sum_{j=0}^{\infty} \psi_j^3}{\left(\sum_{j=0}^{\infty} \psi_j^2\right)^{3/2}}.$$

(c) Using the fourth moment and Theorem 1(c), the kurtosis of X_t is

$$K^{(X)} = \frac{E\left[(X_t-\mu)^4\right]}{\left[E(X_t-\mu)^2\right]^2} = \frac{K^{(u)}\sigma_u^4 \sum_{j=0}^{\infty}\psi_j^4 + 3\sigma_u^4\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2 - 3\sigma_u^4\sum_{j=0}^{\infty}\psi_j^4}{\sigma_u^4\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2} = 3 + \frac{\left(K^{(u)}-3\right)\sum_{j=0}^{\infty}\psi_j^4}{\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2}.$$

(d) It is easily shown that

$$S_t = (X_t-\mu)^2 = \sum_{j=0}^{\infty} u_{t-j}^2\psi_j^2 + 2\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}\psi_i\psi_j u_{t-i}u_{t-j}$$

and

$$S_{t+k} = (X_{t+k}-\mu)^2 = \sum_{j=0}^{\infty} u_{t+k-j}^2\psi_j^2 + 2\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}\psi_i\psi_j u_{t+k-i}u_{t+k-j}.$$

Since u_t 's are uncorrelated, we have

$$E(S_tS_{t+k}) = E\left[\left(\sum_{i=0}^{\infty}u_{t-i}^2\psi_i^2\right)\left(\sum_{j=0}^{\infty}u_{t+k-j}^2\psi_j^2\right)\right] + 4E\left[\left(\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}\psi_i\psi_j u_{t-i}u_{t-j}\right)\left(\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}\psi_i\psi_j u_{t+k-i}u_{t+k-j}\right)\right]$$

$$= \left(K^{(u)}-1\right)\sigma_u^4\sum_{j=0}^{\infty}\psi_j^2\psi_{j+k}^2 + \sigma_u^4\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2 - 2\sigma_u^4\left[\left(\sum_{j=0}^{\infty}\psi_j\psi_{j+k}\right)^2 - \sum_{j=0}^{\infty}\psi_j^2\psi_{j+k}^2\right]$$

$$= \left(K^{(u)}-3\right)\sigma_u^4\sum_{j=0}^{\infty}\psi_j^2\psi_{j+k}^2 + \sigma_u^4\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2 + 2\sigma_u^4\left[\left(\sum_{j=0}^{\infty}\psi_j\psi_{j+k}\right)^2\right].$$

Therefore, the covariance is given by

$$\mathrm{Cov}(S_tS_{t+k}) = E(S_tS_{t+k}) - E(S_t)E(S_{t+k})$$

$$= \left(K^{(u)}-3\right)\sigma_u^4\sum_{j=0}^{\infty}\psi_j^2\psi_{j+k}^2 + 2\sigma_u^4\left[\left(\sum_{j=0}^{\infty}\psi_j\psi_{j+k}\right)^2\right].$$

and the correlation is

$$\rho_k^S = \frac{\mathrm{Cov}(S_tS_{t+k})}{\mathrm{Var}(S_t)} = \frac{\left(K^{(u)}-3\right)\sum_{j=0}^{\infty}\psi_j^2\psi_{j+k}^2 + 2\left[\left(\sum_{j=0}^{\infty}\psi_j\psi_{j+k}\right)^2\right]}{\left(K^{(u)}-3\right)\sum_{j=0}^{\infty}\psi_j^4 + 2\left[\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2\right]}.$$