

Efficient estimation in ZIP models with applications to count data

Nurul N. Mohamad, Ibrahim Mohamed*, Ng Kok-Haur, Mohd S. Yahya
*Institute of Mathematical Sciences, Faculty of Science, University of Malaya,
50603 Kuala Lumpur, Malaysia.*

* Corresponding author: imohamed@um.edu.my

Abstract

Estimating functions have been used in estimating parameters of many continuous time series models. However, this method has not been applied to models involving count data. In this paper, we use quadratic estimating functions (QEF) to derive estimators for the joint estimation of the conditional mean and variance parameters of count data models, specifically the basic zero-inflated Poisson (ZIP) model, ZIP regression model and integer-valued generalized autoregressive heteroscedastic model with ZIP conditional distribution. Results show that the estimators derived from QEF method, which uses information from combined estimating functions, is more informative than linear estimating functions (LEF) method that only uses information from component estimating functions. Finally, we also fit the real data sets using the ZIP models via QEF, LEF and maximum likelihood methods, and in so doing, demonstrate the superiority of the QEF method in practice.

Keywords: Count data; information matrix; linear estimating functions; quadratic estimating functions; zero-inflated Poisson

1. Introduction

Count data are frequently encountered in many biomedical, epidemiological, industrial and public health applications. In practice, especially in the medical field, many count data sets have a high frequency of zeroes. For example, for diseases with low infection rates, the observed counts typically contains a large number of zeroes, although the counts can also be very large during the outbreak period. For such data set, Lambert (1992) introduced a zero-inflated Poisson (ZIP) regression model. His study showed that the ZIP model is not only easy to interpret, but also leads to more refined data analysis as it can accommodate overdispersion.

Following the findings, many studies and applications of the ZIP model have been conducted. For instance, Baksh *et al.* (2011) proposed the overdispersion test for the ZIP model, while Zhu (2012) introduced the model inspired by the generalized autoregressive conditional heteroscedastic (GARCH) model. In the GARCH model, the integer-valued case with a conditional distribution has ZIP distribution instead of normal distribution. It is

denoted as denoted as ZIPINGARCH (p, q) , where p and q are positive integers.

The maximum likelihood (ML) method is commonly used for estimating the parameters of ZIP models when the distribution of the data is known. However, the method does not always perform well under certain circumstances, see for example, Bahadur (1958), Crowder (1987) and Vinod (1997). Furthermore, as pointed out by Nanjundan and Naika (2012), the ML estimator of ZIP models does not have closed-form expression. Therefore, various estimation methods have been proposed as an alternative to the ML method. Some of these are a recursive technique based on the two-step least squares estimator (Abaza, 1982), Monte Carlo EM method (Chan and Ledolfer, 1995), method of moments (Kharrati-Kopaei and Faghih, 2011) and quasi-likelihood (Staub and Winkelmann, 2012).

The semiparametric approach based on the theory of estimating functions (EF) (Godambe, 1985) has been proposed to estimate the parameters of time series models. The EF approach uses the information based on

the first two conditional moments, which is known as linear estimating functions (LEF). This method has been successfully applied in continuous time series models, see Thavaneswaran and Abraham (1988), Chandra and Taniguchi (2001), Bera *et al.* (2006), Merkouris (2007), Allen *et al.* (2013) and Thavaneswaran *et al.* (2012 & 2015). Meanwhile, the advantages of the LEF method have been highlighted in many papers, Bera *et al.* (2006) stated that the LEF approach is a sufficiently flexible moment-based estimation method. It is very useful in econometric applications. In addition, Godambe and Heyde (2010) showed that the LEF estimator yields asymptotically the shortest confidence interval. Moreover, Ng and Peiris (2013) found that the LEF method is more computationally efficient and easy to apply in practice than the ML method. They also argued that the LEF method is easier to evaluate, and estimates can be obtained without the requirement of the assumption of the distribution of errors.

Liang *et al.* (2011) extended the LEF method to the quadratic estimating functions (QEF) method, which involves the first four conditional moments. Their results showed that the QEF method is more informative than the LEF method and gives lower standard errors of the estimated parameters. Furthermore, Thavaneswaran *et al.* (2015) showed that, this extension leads to an improvement in terms of the efficiency of resulting estimate. It has standard asymptotic properties, such as consistency and asymptotic normality. The QEF method also removes the problem of identifiability. On top of that, the Monte Carlo simulation results presented in Ng *et al.* (2015) also showed that the QEF estimators outperform the LEF estimators in almost all cases when applied on the autoregressive conditional duration model.

To our knowledge, the QEF method has never been applied to count data models. Therefore, it is in our interest to investigate the performance of the QEF method as an alternative method in the parameter estimation of these models. We have derived the opti-

mal estimating functions of QEF for the three types of ZIP models, namely the basic ZIP, ZIP regression and ZIPINGARCH(p, q) time series models. We also obtained the information for these ZIP models and then compared them with that from the LEF method. Concurrently, the QEF, LEF and ML methods were applied into real data sets to estimate the model parameters together with their respective standard errors. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) values were calculated to determine the best fitted model.

This paper is organized as follows: Section 2 discusses the theoretical basis for the QEF and LEF methods. In Section 3, we use the QEF method to derive the optimal estimating functions and the information for the three ZIP models. The applicability of the QEF method on empirical examples is presented in Section 4. Finally, concluding remarks are given in Section 5.

2. Parameter estimation methods

This section discusses the LEF and QEF estimation methods.

2.1 Linear estimating functions

Godambe (1985) introduced the estimating functions approach to estimate the parameters of linear and non-linear time series models.

Let $\{y_t\}$ be a discrete time series process depending on a vector parameter θ that belongs to an open subset Θ of the p -dimensional Euclidean space. Let \mathfrak{S}_{t-1}^y be the σ -field generated by $\{y_1, y_2, \dots, y_{t-1}\}$ for $t \geq 1$. Consider a q -dimensional vector $\mathbf{h}_t = \mathbf{h}_t(y_1, y_2, \dots, y_{t-1}, \theta)$ for $1 \leq t \leq n$ which is a martingale difference and let \mathbf{a}_{t-1} be $p \times q$ matrices depending on $\{y_1, y_2, \dots, y_{t-1}\}$. Let \mathfrak{M} be the set of p -dimensional estimating functions $\mathbf{g}_n(\theta)$

$$\mathfrak{M} = \{\mathbf{g}_h(\theta) : \mathbf{g}_h(\theta) = \sum_{t=1}^n \mathbf{a}_{t-1} \mathbf{h}_t\}. \quad (1)$$

An estimate of θ can be obtained by solving the equation $\mathbf{g}_h(\theta) = \mathbf{0}$. Godambe (1985) assumed that the estimating functions $\mathbf{g}_h(\theta)$

are almost surely differentiable with respect to the components of θ so that $E\left[\frac{\partial \mathbf{h}_t}{\partial \theta} | \mathfrak{S}_{t-1}^y\right]$ and $E\left[\mathbf{g}_h(\theta) \mathbf{g}_h'(\theta) | \mathfrak{S}_{t-1}^y\right]$ are nonsingular for all θ for each $t \geq 1$. Moreover, the $p \times p$ matrix $E\left[\mathbf{g}_h(\theta) \mathbf{g}_h'(\theta) | \mathfrak{S}_{t-1}^y\right]$ is assumed to be positive for all θ . Therefore, the optimal $\mathbf{g}_h^*(\theta)$ is given by

$$\begin{aligned} \mathbf{g}_h^*(\theta) &= \sum_{t=1}^n \mathbf{a}_{t-1}^* \mathbf{h}_t \\ &= \sum_{t=1}^n \left(E\left[\frac{\partial \mathbf{h}_t}{\partial \theta} | \mathfrak{S}_{t-1}^y\right] \right)' \\ &\quad \times \left(E\left\{ \mathbf{h}_t \mathbf{h}_t' | \mathfrak{S}_{t-1}^y \right\} \right)^{-1} \mathbf{h}_t, \end{aligned}$$

and the corresponding optimal information matrix is

$$\begin{aligned} \mathbf{I}_{\mathbf{g}_h^*}(\theta) &= \sum_{t=1}^n \left(E\left[\frac{\partial \mathbf{h}_t}{\partial \theta} | \mathfrak{S}_{t-1}^y\right] \right)' \\ &\quad \times \left(E\left[\mathbf{h}_t \mathbf{h}_t' | \mathfrak{S}_{t-1}^y \right] \right)^{-1} \\ &\quad \times \left(E\left[\frac{\partial \mathbf{h}_t}{\partial \theta} | \mathfrak{S}_{t-1}^y\right] \right). \end{aligned}$$

2.2. Quadratic estimating functions

Here we assume that the discrete time stochastic process $\{y_t\}$ has the following conditional moments that depend only on the parameters θ . These are

$$\begin{aligned} \mu_t(\theta) &= E[y_t | \mathfrak{S}_{t-1}^y], \\ \sigma_t^2(\theta) &= E[(y_t - \mu_t(\theta))^2 | \mathfrak{S}_{t-1}^y], \\ \gamma_t(\theta) &= \frac{1}{\sigma_t^3(\theta)} E[(y_t - \mu_t(\theta))^3 | \mathfrak{S}_{t-1}^y], \\ \kappa_t(\theta) &= \frac{1}{\sigma_t^4(\theta)} E[(y_t - \mu_t(\theta))^4 | \mathfrak{S}_{t-1}^y] - 3. \end{aligned}$$

We intend to estimate the parameter θ using two classes of martingale differences

$$\{m_t(\theta) = y_t - \mu_t(\theta), t = 1, 2, \dots, n\}, \quad (2)$$

$$\{s_t(\theta) = m_t^2(\theta) - \sigma_t^2(\theta), t = 1, 2, \dots, n\}, \quad (3)$$

such that

$$\begin{aligned} \langle m \rangle_t &= E[m_t^2(\theta) | \mathfrak{S}_{t-1}^y] \\ &= \sigma_t^2(\theta), \\ \langle s \rangle_t &= E[s_t^2(\theta) | \mathfrak{S}_{t-1}^y] \\ &= \sigma_t^4(\theta)(\kappa_t(\theta) + 2), \\ \langle m, s \rangle_t &= E[m_t(\theta) s_t(\theta) | \mathfrak{S}_{t-1}^y] \\ &= \sigma_t^3(\theta) \gamma_t(\theta). \end{aligned}$$

The optimal estimating functions based on the martingale differences $m_t(\theta)$ and $s_t(\theta)$ are

$$\begin{aligned} \mathbf{g}_m^*(\theta) &= - \sum_{t=1}^n \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{m_t(\theta)}{\langle m \rangle_t}, \\ \mathbf{g}_s^*(\theta) &= - \sum_{t=1}^n \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{s_t(\theta)}{\langle s \rangle_t}, \end{aligned}$$

respectively. The information associated with $\mathbf{g}_m^*(\theta)$ and $\mathbf{g}_s^*(\theta)$ are given as

$$\begin{aligned} \mathbf{I}_{\mathbf{g}_m^*}(\theta) &= \sum_{t=1}^n \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} \frac{1}{\langle m \rangle_t}, \\ \mathbf{I}_{\mathbf{g}_s^*}(\theta) &= \sum_{t=1}^n \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} \frac{1}{\langle s \rangle_t}, \end{aligned}$$

respectively. The optimal QEF and its corresponding information matrix are given by Liang *et al.* (2011) in Theorem 1.

Theorem 1: In the class of all QEF of the form

$$G_Q = \left\{ \mathbf{g}_Q(\theta) = \sum_{t=1}^n (\mathbf{a}_{t-1} m_t(\theta) + \mathbf{b}_{t-1} s_t(\theta)) \right\},$$

(a) the optimal estimating function is given by

$$\mathbf{g}_Q^*(\theta) = \sum_{t=1}^n (\mathbf{a}_{t-1}^* m_t(\theta) + \mathbf{b}_{t-1}^* s_t(\theta)),$$

where $\mathbf{a}_{t-1}^* = R_t Q_t$ and $\mathbf{b}_{t-1}^* = R_t \mathfrak{M}_t$;

(b) the information matrix $\mathbf{I}_{\mathbf{g}_Q^*}(\theta)$ is

$$\mathbf{I}_{\mathbf{g}_Q^*}(\theta) = \sum_{t=1}^n R_t \left(\frac{N_t}{\langle m \rangle_t} + \frac{V_t}{\langle s \rangle_t} - K_t \right);$$

(c) the gain in information $\mathbf{I}_{\mathbf{g}_Q^*}(\theta) - \mathbf{I}_{\mathbf{g}_m^*}(\theta)$ is

$$\begin{aligned} \mathbf{I}_{\mathbf{g}_Q^*}(\theta) - \mathbf{I}_{\mathbf{g}_m^*}(\theta) &= \sum_{t=1}^n R_t \left(N_t \frac{\langle m, s \rangle_t^2}{\langle m \rangle_t^2 \langle s \rangle_t} + \frac{V_t}{\langle s \rangle_t} - K_t \right); \end{aligned}$$

(d) the gain in information $\mathbf{I}_{\mathbf{g}_Q^*}(\theta) - \mathbf{I}_{\mathbf{g}_s^*}(\theta)$ is

$$\begin{aligned} \mathbf{I}_{\mathbf{g}_Q^*}(\theta) - \mathbf{I}_{\mathbf{g}_s^*}(\theta) &= \sum_{t=1}^n R_t \left(\frac{N_t}{\langle m \rangle_t} + V_t R_t \frac{\langle m, s \rangle_t^2}{\langle m \rangle_t \langle s \rangle_t} - K_t \right), \end{aligned}$$

$$\begin{aligned} \text{where } R_t &= \left(1 - \frac{\langle m, s \rangle_t^2}{\langle m \rangle_t \langle s \rangle_t}\right)^{-1}, \\ Q_t &= \left(-\frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{\langle m \rangle_t} + \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\langle m, s \rangle_t}{\langle m \rangle_t \langle s \rangle_t}\right), \\ \mathfrak{M}_t &= \left(\frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\langle m, s \rangle_t}{\langle m \rangle_t \langle s \rangle_t} - \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{\langle s \rangle_t}\right), \\ N_t &= \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}, V_t = \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \text{ and} \\ K_t &= J_t \left(\frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right) \\ \text{with } \mathfrak{J}_t &= \frac{\langle m, s \rangle_t}{\langle m \rangle_t \langle s \rangle_t}. \end{aligned}$$

3. Zero-inflated models

There are three types of ZIP models considered: basic ZIP, ZIP regression and ZIPINGARCH(p, q) time series models. Let us define the probability mass function (pmf) of a zero inflated count data model as

$$f(y) = \begin{cases} \omega + (1 - \omega)g(y) & \text{for } y = 0, \\ (1 - \omega)g(y) & \text{for } y = 1, 2, 3, \dots, \end{cases} \quad (4)$$

where y is a count-valued random variable, $\omega \in [0, 1]$ is a zero-inflation parameter (the probability of a strategic zero), and $g(\cdot)$ is the probability function of the *parent* count model.

The mean of the zero-inflated count data model is

$$E(y) = \sum_{y=0}^{\infty} y f(y) = (1 - \omega)E_g(y),$$

where $E_g(y)$ denotes the mean of the parent distribution. A full parametric zero-inflated count data model is obtained once the probability function of the parent count model is specified. For the next three subsections, we will illustrate the three ZIP models.

3.1 Basic ZIP model

The pmf for this model can be obtained from equation (4) by letting

$$g(y; \lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}, \quad \lambda > 0,$$

with mean $E_g(y) = \lambda$, $\mu_1(\boldsymbol{\theta}) = E[y_t | \mathfrak{S}_{t-1}^y] = (1 - \omega)\lambda$ and $\mu_2(\boldsymbol{\theta}) = E[y_t^2 | \mathfrak{S}_{t-1}^y] = \lambda(1 -$

$\omega)(\lambda + 1)$. The parameter of interest for this model is $\boldsymbol{\theta} = (\lambda, \omega)'$. Following Kharrati-Kopaei and Faghih (2011), we define two martingale differences, $m_t(\boldsymbol{\theta}) = y_t - \mu_1(\boldsymbol{\theta})$ and $S_t(\boldsymbol{\theta}) = y_t^2 - \mu_2(\boldsymbol{\theta})$, respectively. Using the results, the elements of variance-covariance of martingale differences are defined as

$$\begin{aligned} \sigma_{11} &= Var[y_t | \mathfrak{S}_{t-1}^y] \\ &= \mu_1(\boldsymbol{\theta})(1 + \lambda) - (\mu_1(\boldsymbol{\theta}))^2, \\ \sigma_{12} &= Cov[y_t y_t^2 | \mathfrak{S}_{t-1}^y] \\ &= \mu_1(\boldsymbol{\theta})(\lambda^2 + 3\lambda + 1 - \mu_1(\boldsymbol{\theta})(1 + \lambda)), \\ \sigma_{22} &= Var[y_t^2 | \mathfrak{S}_{t-1}^y] \\ &= \mu_1(\boldsymbol{\theta})(\lambda^3 + 6\lambda^2 + 7\lambda + 1 - \mu_1(\boldsymbol{\theta})(1 + \lambda)^2). \end{aligned}$$

It is easily shown that $\langle m \rangle_t = \sigma_{11}$, $\langle S \rangle_t = \sigma_{22}$ and $\langle m, S \rangle_t = \sigma_{12}$. The derivatives of $\mu_1(\boldsymbol{\theta})$ and $\mu_2(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ are

$$\begin{aligned} \frac{\partial \mu_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= (1 - \omega, -\lambda)', \\ \frac{\partial \mu_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= ((1 - \omega)(1 + 2\lambda), -(\lambda + \lambda^2))'. \end{aligned}$$

From Theorem 1, the optimal QEF for each parameter λ and ω are given by

$$\begin{aligned} \mathbf{g}_Q^*(\lambda) &= (1 - \omega) \left(\frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \right) \sum_{t=1}^n J_{1,t}, \\ \mathbf{g}_Q^*(\omega) &= \lambda \left(\frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \right) \sum_{t=1}^n J_{2,t}, \end{aligned}$$

respectively, where

$$\begin{aligned} J_{1,t} &= \begin{pmatrix} -\sigma_{22} + \sigma_{12}H_{1,t} \\ \sigma_{12} - \sigma_{11}H_{1,t} \end{pmatrix} m_t(\boldsymbol{\theta}) + \\ &\quad \begin{pmatrix} \sigma_{12} - \sigma_{11}H_{1,t} \\ \sigma_{22} - H_{2,t}\sigma_{12} \end{pmatrix} S_t(\boldsymbol{\theta}), \\ J_{2,t} &= \begin{pmatrix} \sigma_{22} - H_{2,t}\sigma_{12} \\ \sigma_{11}H_{2,t} - \sigma_{12} \end{pmatrix} m_t(\boldsymbol{\theta}) + \\ &\quad \begin{pmatrix} \sigma_{11}H_{2,t} - \sigma_{12} \\ \sigma_{22} - H_{2,t}\sigma_{12} \end{pmatrix} S_t(\boldsymbol{\theta}), \end{aligned}$$

with $H_{1,t} = 1 + 2\lambda$ and $H_{2,t} = 1 + \lambda$.

The corresponding information matrix of $\boldsymbol{\theta}$ is

$$\mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta}) = \begin{bmatrix} I_{\lambda\lambda}^Q & I_{\lambda\omega}^Q \\ I_{\omega\lambda}^Q & I_{\omega\omega}^Q \end{bmatrix},$$

where $I_{\lambda\lambda}^Q = \frac{n(1 - \omega)^2 E_t}{F_t}$, $I_{\omega\omega}^Q = \frac{n\lambda^2 L_t}{F_t}$, and symmetrical elements, $I_{\omega\lambda}^Q = I_{\lambda\omega}^Q =$

$\frac{n\lambda(1-\omega)P_t}{F_t}$ such that $E_t = \sigma_{22} + H_{1,t}(H_{1,t}\sigma_{11} - 2\sigma_{12}), F_t = \sigma_{11}\sigma_{22} - \sigma_{12}^2, L_t = \sigma_{22} + \sigma_{11}H_{2,t}^2 - 2H_{2,t}\sigma_{12}, P_t = -\sigma_{22} - \sigma_{11}(1 + 3\lambda + 2\lambda^2) + \sigma_{12}(2 + 3\lambda)$.

For simplicity, we only compare the information for the parameter λ . Hence, the information gain by estimating functions for martingale differences, $m_t(\boldsymbol{\theta})$ and $S_t(\boldsymbol{\theta})$ are $I_{\lambda\lambda}^m = n(1-\omega)^2/\sigma_{11}$ and $I_{\lambda\lambda}^S = n(1-\omega)^2(1+2\lambda)/\sigma_{11}$, respectively. It is noted that the value of the denominator of $I_{\lambda\lambda}^Q$ is small when the numerator is large, leading to a greater value of information gained when compared to $I_{\lambda\lambda}^m$ and $I_{\lambda\lambda}^S$, i.e., $I_{\lambda\lambda}^Q > I_{\lambda\lambda}^m$ and $I_{\lambda\lambda}^Q > I_{\lambda\lambda}^S$. Hence, one can say that, the combined estimating functions is more informative than the individual elements.

3.2 ZIP regression model

We consider the ZIP regression model. In some cases, we may parameterize both λ and ω in terms of exogenous explanatory variables, say x and z . Following the definition given by Staub and Winkelmann (2012), we assume that $\lambda = \exp(\lambda_0 + \lambda_1 x)$ and $\omega = \frac{\exp(\delta_0 + \delta_1 z)}{1 + \exp(\delta_0 + \delta_1 z)}$, where z can be identical to x , overlap with x , or be completely distinct from x .

The parameter is $\boldsymbol{\theta} = (\lambda_0, \lambda_1, \delta_0, \delta_1)'$ and the conditional expectation function of the corresponding ZIP model is given by

$$E(y | x, z) = \frac{\exp(\lambda_0 + \lambda_1 x)}{1 + \exp(\delta_0 + \delta_1 z)}. \quad (5)$$

Here, we let independent counts to be y_t , where $t = 1, 2, \dots, n$, with λ_t and ω_t coming from λ and ω as mentioned above. Hence, the conditional mean, variance, skewness and kurtosis are defined as

$$\begin{aligned} \mu_t(\boldsymbol{\theta}) &= \frac{\exp(\lambda_0 + \lambda_1 x_t)}{1 + \exp(\delta_0 + \delta_1 z_t)}, \\ \sigma_t^2(\boldsymbol{\theta}) &= \frac{\exp(\lambda_0 + \lambda_1 x_t)T_t}{[1 + \exp(\delta_0 + \delta_1 z_t)]^2}, \end{aligned}$$

$$\begin{aligned} \gamma_t(\boldsymbol{\theta}) &= \frac{1 + 3\lambda_t\omega_t + \lambda_t^2\omega_t + 2\lambda_t^2\omega_t^2}{\mu_t [1 + \lambda_t\omega_t]^{\frac{3}{2}}}, \\ \kappa_t(\boldsymbol{\theta}) &= \frac{F_t}{(1 - \omega_t)\lambda_t(1 + \omega_t\lambda_t)^2}, \end{aligned}$$

respectively, where $F_t = \omega_t\lambda_t^3(6\omega_t^2 - 6\omega_t + 1) + 6\omega_t\lambda_t^2(2\omega_t - 1) + 7\omega_t\lambda_t + 1$ and $T_t = 1 + \exp(\delta_0 + \delta_1 z_t) + \exp(\lambda_0 + \lambda_1 x_t + \delta_0 + \delta_1 z_t)$.

The derivative of $\mu_t(\boldsymbol{\theta})$ with respect to each parameter is $\partial\mu_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = (B_1, B_2, B_3, B_4)'$, where

$$\begin{aligned} B_{1,t} &= \frac{\exp(\lambda_0 + \lambda_1 x_t)}{1 + \exp(\delta_0 + \delta_1 z_t)}, \\ B_{2,t} &= \frac{\exp(\lambda_0 + \lambda_1 x_t)x_t}{1 + \exp(\delta_0 + \delta_1 z_t)}, \\ B_{3,t} &= -\frac{\exp(\lambda_0 + \lambda_1 x_t + \delta_0 + \delta_1 z_t)}{[1 + \exp(\delta_0 + \delta_1 z_t)]^2}, \\ B_{4,t} &= -\frac{\exp(\lambda_0 + \lambda_1 x_t + \delta_0 + \delta_1 z_t)z_t}{[1 + \exp(\delta_0 + \delta_1 z_t)]^2}. \end{aligned}$$

Now, let $\Upsilon_t = 1/(\eta_t\sigma_t^4(\boldsymbol{\theta}))\mathfrak{D}_t$, where $\eta_t = \kappa_t(\boldsymbol{\theta}) + 2 - \gamma_t^2(\boldsymbol{\theta})$, and $\mathfrak{D}_t = -\sigma_t^2(\boldsymbol{\theta})(\kappa_t(\boldsymbol{\theta}) + 2 - (1 + \lambda_t\omega_t)) + \gamma_t\sigma_t(\boldsymbol{\theta})(Y_t - \mu_t(\boldsymbol{\theta}))(1 + \lambda_t\omega_t + (Y_t - \mu_t(\boldsymbol{\theta}))) - \sigma_t^3(\boldsymbol{\theta})\gamma_t(\boldsymbol{\theta})$ with $t = 1, 2, \dots, n$.

The optimal QEF for $\lambda_0, \lambda_1, \delta_0$ and δ_1 are respectively

$$\begin{aligned} \mathbf{g}_Q^*(\lambda_0) &= \sum_{t=1}^n \Upsilon_t B_{1,t}, \quad \mathbf{g}_Q^*(\lambda_1) = \sum_{t=1}^n \Upsilon_t B_{2,t}, \\ \mathbf{g}_Q^*(\delta_0) &= \sum_{t=1}^n \Upsilon_t B_{3,t}, \quad \mathbf{g}_Q^*(\delta_1) = \sum_{t=1}^n \Upsilon_t B_{4,t}. \end{aligned}$$

The corresponding information matrix for $\boldsymbol{\theta}$ based on the QEF method is therefore

$$\begin{aligned} \mathbf{I}_{\mathbf{g}_Q^*}(\boldsymbol{\theta}) &= \sum_{i=1}^n \left(\frac{1}{\sigma_t^3(\boldsymbol{\theta})\eta_t} \right) \\ &\quad \times \begin{pmatrix} \sigma_t(\boldsymbol{\theta})(\kappa_t(\boldsymbol{\theta}) + 2)W_t^2 \\ -2\gamma_t(\boldsymbol{\theta})W_t \end{pmatrix} \mathfrak{z}_t, \quad (6) \end{aligned}$$

where

$$\mathfrak{z}_t = \begin{bmatrix} B_{1,t}^2 & B_{1,t}B_{2,t} & B_{1,t}B_{3,t} & B_{1,t}B_{4,t} \\ B_{2,t}B_{1,t} & B_{2,t}^2 & B_{2,t}B_{3,t} & B_{2,t}B_{4,t} \\ B_{3,t}B_{1,t} & B_{3,t}B_{2,t} & B_{3,t}^2 & B_{3,t}B_{4,t} \\ B_{4,t}B_{1,t} & B_{4,t}B_{2,t} & B_{4,t}B_{3,t} & B_{4,t}^2 \end{bmatrix}$$

$$\text{and } W_t = \frac{\exp(\lambda_0 + \lambda_1 x_t + \delta_0 + \delta_1 z_t)}{\exp(\delta_0 + \delta_1 z_t)}.$$

For the purpose of comparison, we focus only on the parameters λ_0 and λ_1 with the information gained by the use of estimating functions based on the single element of martingale differences, $m_t(\boldsymbol{\theta}) = y_t - \mu_t(\boldsymbol{\theta})$ and $s_t(\boldsymbol{\theta}) = m_t^2(\boldsymbol{\theta}) - \sigma_t^2(\boldsymbol{\theta})$. For $m_t(\boldsymbol{\theta})$, the information matrices based on the parameter λ are

$$I_{\lambda_0\lambda_0}^m = \sum_{i=1}^n B_{1,t}^2 / \sigma_t^2(\boldsymbol{\theta}),$$

and

$$I_{\lambda_1\lambda_1}^m = \sum_{i=1}^n B_{2,t}^2 / \sigma_t^2(\boldsymbol{\theta}).$$

While for $s_t(\boldsymbol{\theta})$, we have

$$I_{\lambda_0\lambda_0}^s = \sum_{i=1}^n \bar{h}_t W_t^2 B_{1,t}^2,$$

and

$$I_{\lambda_1\lambda_1}^s = \sum_{i=1}^n l \bar{h}_t W_t^2 B_{2,t}^2,$$

where $\bar{h}_t = [1/\{\sigma_t^4(\boldsymbol{\theta})(\kappa_t(\boldsymbol{\theta}) + 2)\}]$. Through the derivation of the information matrix in equation (6), we have the following results, where $I_{\lambda_0\lambda_0}^Q > I_{\lambda_0\lambda_0}^m$ and $I_{\lambda_0\lambda_0}^Q > I_{\lambda_0\lambda_0}^s$ as well as $I_{\lambda_1\lambda_1}^Q > I_{\lambda_1\lambda_1}^m$ and $I_{\lambda_1\lambda_1}^Q > I_{\lambda_1\lambda_1}^s$. Therefore, for the ZIP regression model, we can conclude that the information acquired using the QEF method is more informative than that of its linear components.

3.3 ZIPINGARCH(p, q) time series model

In this last part, we focus on the model analogues to the GARCH(p, q) model with its conditional distribution following ZIP. The model is denoted as ZIPINGARCH(p, q). Let y_t denote a count time series with excess zeroes conditional on \mathfrak{S}_{t-1}^y and modeled by

$$p(y_t | \mathfrak{S}_{t-1}^y) = \begin{cases} \omega + (1 - \omega) e^{-\lambda_t(\boldsymbol{\theta})} & \text{for } y_t = 0, \\ (1 - \omega) \frac{e^{-\lambda_t(\boldsymbol{\theta})} [\lambda_t(\boldsymbol{\theta})]^{y_t}}{y_t!} & \text{for } y_t > 0. \end{cases}$$

Here, $\lambda_t(\boldsymbol{\theta})$ is the intensity parameter based on the baseline Poisson distribution and ω is the zero-inflated parameter with $\lambda_t(\boldsymbol{\theta})$ defined by Zhu (2012) as

$$\lambda_t(\boldsymbol{\theta}) = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{j=1}^q \beta_j \lambda_{t-j}(\boldsymbol{\theta}).$$

Suppose that the observations $Y = (y_1, y_2, \dots, y_n)$ are generated from the model. Then, $\boldsymbol{\theta} = (\omega, \alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q)'$. The mean, variance, skewness and kurtosis of y_t conditional on \mathfrak{S}_{t-1}^y are

$$\begin{aligned} \mu_t(\boldsymbol{\theta}) &= (1 - \omega)\lambda_t(\boldsymbol{\theta}), \\ \sigma_t^2(\boldsymbol{\theta}) &= (1 - \omega)\lambda_t(\boldsymbol{\theta})(1 + \omega\lambda_t(\boldsymbol{\theta})), \\ \gamma_t(\boldsymbol{\theta}) &= \frac{\omega(2\omega + 1) [\lambda_t(\boldsymbol{\theta})]^2 + 3\omega\lambda_t(\boldsymbol{\theta}) + 1}{[(1 - \omega)\lambda_t(\boldsymbol{\theta})]^{1/2} [1 + \omega\lambda_t(\boldsymbol{\theta})]^{3/2}}, \end{aligned}$$

$$\kappa_t(\boldsymbol{\theta}) = \frac{\mathfrak{T}_t}{(1 - \omega)\lambda_t(\boldsymbol{\theta}) [1 + \omega\lambda_t(\boldsymbol{\theta})]^2},$$

respectively, where $\mathfrak{T}_t = \omega(6\omega^2 - 6\omega + 1) [\lambda_t(\boldsymbol{\theta})]^3 + 6\omega(2\omega - 1) [\lambda_t(\boldsymbol{\theta})]^2 + 7\omega\lambda_t(\boldsymbol{\theta}) + 1$.

By taking $m_t(\boldsymbol{\theta}) = y_t - \mu_t(\boldsymbol{\theta})$ and $s_t(\boldsymbol{\theta}) = m_t^2(\boldsymbol{\theta}) - \sigma_t^2(\boldsymbol{\theta})$, we have $\langle m \rangle_t = \sigma_t^2(\boldsymbol{\theta})$, $\langle s \rangle_t = \sigma_t^4(\boldsymbol{\theta})(\kappa_t(\boldsymbol{\theta}) + 2)$ and $\langle m, s \rangle_t = \sigma_t^3(\boldsymbol{\theta})\gamma_t(\boldsymbol{\theta})$. For the derivatives of the $\mu_t(\boldsymbol{\theta})$ and $\sigma_t^2(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, we have

$$\begin{aligned} \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \begin{pmatrix} -\lambda_t(\boldsymbol{\theta}), (1 - \omega) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \alpha_0}, \dots, \\ (1 - \omega) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \alpha_p}, (1 - \omega) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \beta_1}, \dots, \\ (1 - \omega) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \beta_q} \end{pmatrix}' \\ &= \begin{pmatrix} A_{1,t}, A_{2,t}, A_{(3,1),t}, \dots, \\ A_{(3,p),t}, \dots, A_{(4,1),t}, \dots, A_{(4,q),t} \end{pmatrix}', \\ \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \begin{pmatrix} -\lambda_t(\boldsymbol{\theta})(1 + 2\omega\lambda_t(\boldsymbol{\theta})), C_t \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \alpha_0}, \\ \dots, C_t \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \alpha_p}, \\ C_t \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \beta_1}, \dots, C_t \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \beta_q} \end{pmatrix}' \\ &= \begin{pmatrix} S_{1,t}, S_{2,t}, S_{(3,1),t}, \dots, S_{(3,p),t}, \\ \dots, S_{(4,1),t}, \dots, S_{(4,q),t} \end{pmatrix}', \end{aligned}$$

where $C_t = (1 - \omega)(1 + 2\omega\lambda_t(\boldsymbol{\theta}))$. Now, let $A_{k,t}^m = \frac{-A_{k,t}}{\langle m \rangle_t}$, $S_{k,t}^m = \frac{S_{k,t} \langle m, s \rangle_t}{\langle m \rangle_t \langle s \rangle_t}$, $P_{k,t}^v = \frac{A_{k,t} \langle m, s \rangle_t}{\langle m \rangle_t \langle s \rangle_t}$, and $Q_{k,t}^v = \frac{-S_{k,t}}{\langle s \rangle_t}$ and $B_t = \left(1 - \frac{\langle m, s \rangle_t^2}{\langle m \rangle_t \langle s \rangle_t}\right)^{-1}$.

From Theorem 1(a), \mathbf{a}_{t-1}^* and \mathbf{b}_{t-1}^* are given in matrix form by

$$\mathbf{a}_{t-1}^* = B_t \begin{pmatrix} A_{1,t}^m + S_{1,t}^m, \dots, A_{(3,1),t}^m + S_{(3,1),t}^m, \\ \dots, A_{(3,p),t}^m + S_{(3,p),t}^m, A_{(4,1),t}^m + S_{(4,1),t}^m, \\ S_{(4,1),t}^m, \dots, A_{(4,q),t}^m + S_{(4,q),t}^m \end{pmatrix},$$

$$\mathbf{b}_{t-1}^* = B_t \begin{pmatrix} P_{1,t}^v + Q_{1,t}^v, \dots, P_{(3,1),t}^v + Q_{(3,1),t}^v, \\ \dots, P_{(3,p),t}^v + Q_{(3,p),t}^v, P_{(4,1),t}^v + Q_{(4,1),t}^v, \\ Q_{(4,1),t}^v, \dots, P_{(4,q),t}^v + Q_{(4,q),t}^v \end{pmatrix}.$$

Thus, the optimal estimating functions for each component of $\boldsymbol{\theta}$ are

$$\mathbf{g}_Q^*(\omega) = \sum_{t=1}^n \begin{pmatrix} (A_{1,t}^m + S_{1,t}^m) m_t(\boldsymbol{\theta}) \\ + (P_{1,t}^v + Q_{1,t}^v) s_t(\boldsymbol{\theta}) \end{pmatrix},$$

$$\mathbf{g}_Q^*(\alpha_0) = \sum_{t=1}^n \begin{pmatrix} (A_{2,t}^m + S_{2,t}^m) m_t(\boldsymbol{\theta}) \\ + (P_{2,t}^v + Q_{2,t}^v) s_t(\boldsymbol{\theta}) \end{pmatrix},$$

$$\mathbf{g}_Q^*(\alpha_i) = \sum_{t=1}^n \begin{pmatrix} (A_{(3,i),t}^m + S_{(3,i),t}^m) m_t(\boldsymbol{\theta}) \\ + (P_{(3,i),t}^v + Q_{(3,i),t}^v) s_t(\boldsymbol{\theta}) \end{pmatrix},$$

for $i = 1, 2, \dots, p$, and

$$\mathbf{g}_Q^*(\beta_j) = \sum_{t=1}^n \begin{pmatrix} (A_{(4,j),t}^m + S_{(4,j),t}^m) m_t(\boldsymbol{\theta}) \\ + (P_{(4,j),t}^v + Q_{(4,j),t}^v) s_t(\boldsymbol{\theta}) \end{pmatrix},$$

for $j = 1, 2, \dots, q$.

The corresponding information matrix of the optimal QEF for $\boldsymbol{\theta}$ is a $(p + q + 2) \times (p + q + 2)$ matrix with the elements are given by

$$I_{\omega\omega} = \sum_{t=1}^n B_t \left[\frac{A_{1,t}^2}{\langle m \rangle_t} + \frac{S_{1,t}^2}{\langle s \rangle_t} - 2A_{1,t}S_{1,t}\xi_t \right],$$

$$I_{\alpha_0\alpha_0} = \sum_{t=1}^n B_t \left[\frac{A_{2,t}^2}{\langle m \rangle_t} + \frac{S_{2,t}^2}{\langle s \rangle_t} - 2A_{2,t}S_{2,t}\xi_t \right],$$

$$I_{\alpha_i\alpha_i} = \sum_{t=1}^n B_t \left[\frac{A_{(3,i),t}^2}{\langle m \rangle_t} + \frac{S_{(3,i),t}^2}{\langle s \rangle_t} - 2A_{(3,i),t}S_{(3,i),t}\xi_t \right],$$

$$I_{\beta_j\beta_j} = \sum_{t=1}^n B_t \left[\frac{A_{(4,j),t}^2}{\langle m \rangle_t} + \frac{S_{(4,j),t}^2}{\langle s \rangle_t} - 2A_{(4,j),t}S_{(4,j),t}\xi_t \right],$$

$$I_{\omega\alpha_0} = \sum_{t=1}^n B_t \left[\frac{A_{1,t}A_{2,t}}{\langle m \rangle_t} + \frac{S_{1,t}S_{2,t}}{\langle s \rangle_t} - \begin{pmatrix} A_{1,t}S_{2,t} \\ S_{1,t}A_{2,t} \end{pmatrix} \xi_t \right],$$

$$I_{\omega\alpha_i} = \sum_{t=1}^n B_t \left[\frac{A_{1,t}A_{(3,i),t}}{\langle m \rangle_t} + \frac{S_{1,t}S_{(3,i),t}}{\langle s \rangle_t} - \begin{pmatrix} A_{1,t}S_{(3,i),t} \\ S_{1,t}A_{(3,i),t} \end{pmatrix} \xi_t \right],$$

$$I_{\omega\beta_j} = \sum_{t=1}^n B_t \left[\frac{A_{1,t}A_{(4,i),t}}{\langle m \rangle_t} + \frac{S_{1,t}S_{(4,i),t}}{\langle s \rangle_t} - \begin{pmatrix} A_{1,t}S_{(4,i),t} \\ S_{1,t}A_{(4,i),t} \end{pmatrix} \xi_t \right],$$

$$I_{\alpha_0\alpha_i} = \sum_{t=1}^n B_t \left[\frac{A_{(3,i),t}A_{2,t}}{\langle m \rangle_t} + \frac{S_{(3,i),t}S_{2,t}}{\langle s \rangle_t} - \begin{pmatrix} A_{2,t}S_{(3,i),t} \\ S_{2,t}A_{(3,i),t} \end{pmatrix} \xi_t \right],$$

$$I_{\alpha_0\beta_j} = \sum_{t=1}^n B_t \left[\frac{A_{(4,j),t}A_{2,t}}{\langle m \rangle_t} + \frac{S_{(4,j),t}S_{2,t}}{\langle s \rangle_t} - \begin{pmatrix} A_{2,t}S_{(4,j),t} \\ S_{2,t}A_{(4,j),t} \end{pmatrix} \xi_t \right],$$

$$I_{\alpha_i\beta_j} = \sum_{t=1}^n B_t \left[\frac{A_{(3,i),t}A_{(4,j),t}}{\langle m \rangle_t} + \frac{A_{(3,j),t}S_{(4,j),t}}{\langle s \rangle_t} - \begin{pmatrix} A_{(3,i),t}S_{(4,j),t} \\ S_{(3,i),t}A_{(4,j),t} \end{pmatrix} \xi_t \right],$$

with the symmetrical elements, $I_{\omega\alpha_0} = I_{\alpha_0\omega}$, $I_{\omega\alpha_i} = I_{\alpha_i\omega}$, $I_{\omega\beta_j} = I_{\beta_j\omega}$, $I_{\alpha_0\alpha_i} = I_{\alpha_i\alpha_0}$, $I_{\alpha_0\beta_j} = I_{\beta_j\alpha_0}$, $I_{\alpha_i\beta_j} = I_{\beta_j\alpha_i}$ and $\xi_t = \frac{\langle m, s \rangle_t}{\langle m \rangle_t \langle s \rangle_t}$.

Again, we compare the information only for parameter $\lambda_t(\boldsymbol{\theta})$. Using estimating functions based on $m_t(\boldsymbol{\theta}) = y_t - \mu_t(\boldsymbol{\theta})$, the information matrix $\mathbf{I}_{\mathbf{g}_m^*}(\boldsymbol{\theta})$ is computed with the elements

$$I_{\alpha_0\alpha_0}^m = \sum_{t=1}^n A_{2,t}^2 / \sigma_t^2(\boldsymbol{\theta}),$$

$$I_{\alpha_i\alpha_i}^m = \sum_{t=1}^n A_{3,t}^2 / \sigma_t^2(\boldsymbol{\theta}),$$

and

$$I_{\beta_j\beta_j}^m = \sum_{t=1}^n A_{4,t}^2 / \sigma_t^2(\boldsymbol{\theta}).$$

Similarly, the information matrix $\mathbf{I}_{\mathbf{g}_s^*}(\boldsymbol{\theta})$ for

component $s_t(\boldsymbol{\theta})$ are

$$I_{\alpha_0\alpha_0}^s = \sum_{t=1}^n D_t S_{2,t}^2,$$

$$I_{\alpha_i\alpha_i}^s = \sum_{t=1}^n D_t S_{3,t}^2,$$

and

$$I_{\beta_j\beta_j}^s = \sum_{t=1}^n D_t S_{4,t}^2$$

where $D_t = [1/(\sigma_t^4(\boldsymbol{\theta})(\kappa_t(\boldsymbol{\theta}) + 2))]$.

From the information using the QEF method and information via its components $m_t(\boldsymbol{\theta})$ and $s_t(\boldsymbol{\theta})$, it is clear that $I_{\alpha_0\alpha_0}^Q > I_{\alpha_0\alpha_0}^m$, $I_{\alpha_0\alpha_0}^Q > I_{\alpha_0\alpha_0}^s$, $I_{\alpha_i\alpha_i}^Q > I_{\alpha_i\alpha_i}^m$, $I_{\alpha_i\alpha_i}^Q > I_{\alpha_i\alpha_i}^s$, $I_{\beta_j\beta_j}^Q > I_{\beta_j\beta_j}^m$, and $I_{\beta_j\beta_j}^Q > I_{\beta_j\beta_j}^s$.

Therefore, we can suggest that the QEF is more informative than the component estimating functions for the ZIPINGARCH(p, q) time series model.

4. Applications

This section discusses the applications of the QEF method on the ZIP models using two real data sets.

4.1 Basic ZIP model

By adopting a count data time series from the Forecasting Principles site at <http://www.forecastingprinciples.com>, the data represent 144 monthly counts of arson in the 13th police car beat in Pittsburgh, Pennsylvania, USA from January 1990 until December 2001. The data have 54 zeroes, i.e. 37.5% of the series. The plot of data is given in Figure 1.

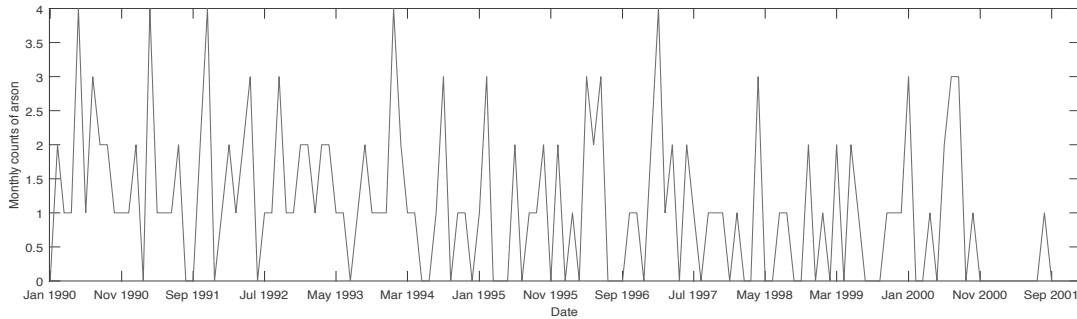


Fig. 1. Monthly counts of arson in the 13th police car beat in Pittsburgh from January 1990 until December 2001

We fitted the basic ZIP model on the data and estimated the parameters using QEF, LEF and ML methods. Using R-CRAN programming language, the ML estimates can be obtained by maximizing the likelihood of the

model through a *nlminb* function, while for QEF and LEF methods, their estimates were obtained by solving their respective simultaneous optimal estimating functions via a *nleqsv* function.

Table 1. Parameter estimates, AIC and BIC for basic ZIP model. Values in parenthesis are standard errors of parameter estimates

Method	$\hat{\lambda}$	$\hat{\omega}$	AIC	BIC
QEF	1.156(0.031)	0.075(0.015)	455.82	621.03
LEF	1.159(0.054)	0.089(0.018)	458.14	627.35
ML	1.158(0.032)	0.083(0.018)	458.68	624.28

Table 1 reports the estimated parameters, estimated standard errors, AIC and BIC for ZIP models using three different estimation methods. We observe that the standard errors of QEF estimates are lower than those of LEF and ML estimates. In addition, the basic ZIP model with QEF estimates gives the lowest AIC and BIC values. This indicates that a model with QEF estimates gives a better model fit for arson data than the model with LEF and ML estimates.

To assess the performance of these QEF estimates to LEF and ML estimates, a simulation study was carried out. We use a block bootstrap method provided by R-CRAN pro-

gramming language: *tseries* package and *tsbootstrap* command with $m=20$ blocks to form a new series and estimate the model parameters of the new series using QEF, LEF and ML methods. The procedure is repeated for $nb=500$ replications and the summary statistics: mean, bias, standard error (SE) and root mean squared error (RMSE) for each parameter are computed. The simulation results are shown in Table 2. We observe that the QEF estimates give lower estimated bias than LEF and ML estimates. In addition, the estimated SEs based on simulation are consistent with the empirical results for all three methods as shown in Table 1.

Table 2. Simulation results based on bootstrap method for basic ZIP model

Method	Estimated parameter	Mean	Bias	SE	RMSE
QEF	$\hat{\lambda}$	1.192	0.037	0.031	0.048
	$\hat{\omega}$	0.079	0.004	0.014	0.015
LEF	$\hat{\lambda}$	1.227	0.068	0.053	0.086
	$\hat{\omega}$	0.079	0.010	0.017	0.020
ML	$\hat{\lambda}$	1.196	0.038	0.033	0.051
	$\hat{\omega}$	0.077	0.006	0.019	0.020

In order to check the model adequacy for the basic ZIP model based on QEF estimates, the Pearson residual is computed which is defined as $\epsilon_t = (y_t - (1 - \hat{\omega})\hat{\lambda}) / \sqrt{\hat{\lambda}(1 - \hat{\omega})(1 + \hat{\lambda})}$. We found that the mean and variance of Pearson residuals are 0.083 and 0.956, respectively. These values are relatively close to zero and unity indicates that the data is adequately fitted to the model. The Ljung-Box (LB) test results

as given in Table3 also indicated that there is no significant serial correlation in the residual. This shows that the data fit well using the basic ZIP model with estimated parameters, $\hat{\lambda}$ and $\hat{\omega}$, say ZIP $(\hat{\lambda}, \hat{\omega})$. Furthermore, Figure 2 shows that the cumulative periodogram plot (see Brockwell and Davis (1991)) does not cross the dotted line. Therefore, we can conclude that, the basic ZIP model via QEF estimates gives a better fit for the arson data.

Table 3. Diagnostics for basic ZIP model

	$LB_{30}(\epsilon_t)$	$LB_{30}(\epsilon_t^2)$
χ^2	23.6	30.4
p-value	0.785	0.477

4.2 ZIP regression model

For the ZIP regression model, we consider data

from the National Medical Expenditure Survey concerning medical care utilization by the

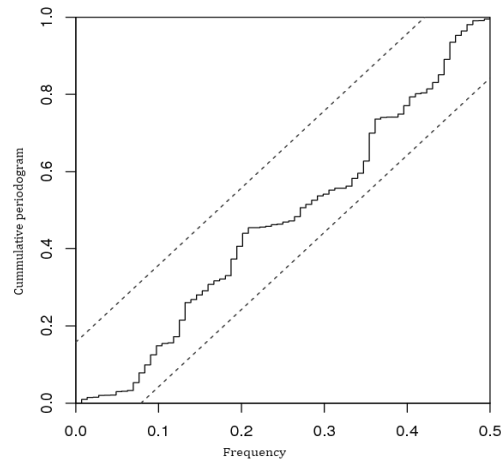


Fig. 2. Cummulative periodogram plot

older American conducted in 1987 and 1988 in the United States. The data set included 4406 people above the age of 66 years that were covered by medicare. The data is available at <http://qed.econ.queensu.ca/jae/1997->

[v12.3/deb-trivedi/](http://qed.econ.queensu.ca/jae/1997-v12.3/deb-trivedi/). The number of patients with chronic conditions (x) and the number of doctor visits in a hospital (y) were the only two variables included from the data (Figure 3).

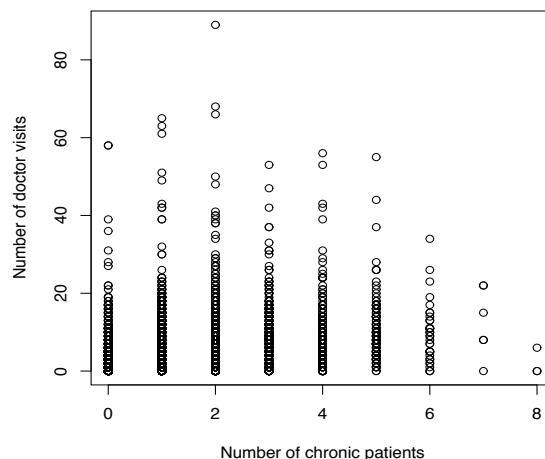


Fig. 3. Number of doctor visits in hospital versus number of patients (≥ 66 years) with chronic conditions

The estimated parameters together with their standard errors using all three methods are shown in Table 4. It is clear that the estimates using the QEF method give the lowest standard errors when compared to the other two methods. Furthermore, the QEF method produces lower AIC and BIC values than LEF and ML methods. This indicates that the ZIP

regression model using QEF estimates gives the best model. For the simulation study, we use the same procedure as described in Section 4.1 (see Table 5). We obtain the same conclusion as discussed in Section 4.1. First, QEF estimates give the smallest bias. Second, the standard errors based on the simulation are consistent with the empirical results in Table 4.

Table 4. Parameter estimates, AIC and BIC for the ZIP regression model. Values in parenthesis are standard errors of parameter estimates

Method	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\delta}_0$	$\hat{\delta}_1$	AIC	BIC
QEF	0.531 (0.010)	0.119 (0.015)	0.125 (0.081)	0.189 (0.013)	1258.36	1425.98
LEF	0.541 (0.014)	0.117 (0.021)	0.129 (0.086)	0.193 (0.015)	1262.23	1428.33
ML	0.532 (0.013)	0.121 (0.018)	0.127 (0.081)	0.190 (0.013)	1260.11	1426.05

Table 5. Simulation results based on bootstrap method for ZIP regression model

Method	Estimated parameter	Mean	Bias	SE	RMSE
QEF	$\hat{\lambda}_0$	0.547	0.016	0.010	0.019
	$\hat{\lambda}_1$	0.122	0.003	0.016	0.017
	$\hat{\delta}_0$	0.183	0.058	0.081	0.100
	$\hat{\delta}_1$	0.188	0.001	0.013	0.014
LEF	$\hat{\lambda}_0$	0.559	0.018	0.015	0.023
	$\hat{\lambda}_1$	0.127	0.010	0.021	0.024
	$\hat{\delta}_0$	0.237	0.108	0.087	0.174
	$\hat{\delta}_1$	0.186	0.007	0.015	0.016
ML	$\hat{\lambda}_0$	0.552	0.020	0.013	0.024
	$\hat{\lambda}_1$	0.129	0.008	0.019	0.021
	$\hat{\delta}_0$	0.246	0.119	0.082	0.145
	$\hat{\delta}_1$	0.188	0.002	0.013	0.015

4.3 ZIPINGARCH(p, q) time series model

For this model, a similar data set as mentioned in Section 4.1 was used. We obtain the parameter estimates, standard errors and AIC and BIC values for ZIPINGARCH (1, 1) model via QEF, LEF and ML estimates (see Table 6). The standard errors of the QEF estimates give comparable results with the ML method, but they are lower than the LEF method. The ZIPINGARCH (1, 1) model using QEF method gives the lowest AIC and BIC values compared to the LEF and ML methods. The same procedure as outlined in Section 4.1 was used. The simulation results can be found

in Table 7. The QEF estimates give lower estimated bias than LEF and ML estimates. We also observe that estimated SEs and RMSEs based on QEF estimates are lower than that LEF and ML estimates. The estimated SEs based on the simulation are agreed in agreement with the empirical results for all three methods as shown in Table 6. The mean and variance of Pearson residuals are close to zero and unity which are 0.039 and 0.992, respectively, indicating the adequacy of the model. The results of the LB test suggest that there is no significant serial correlation in the residuals (Table 8).

Table 6. Parameter estimates, AIC and BIC for ZIPINGARCH(1, 1) model. Values in parenthesis are standard errors of parameter estimates

Method	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\omega}$	AIC	BIC
QEF	0.277 (0.007)	0.089 (0.015)	0.217 (0.049)	0.214 (0.017)	394.16	405.36
LEF	0.281 (0.012)	0.084 (0.023)	0.233 (0.067)	0.208 (0.019)	397.23	408.37
ML	0.279 (0.007)	0.087 (0.014)	0.213 (0.051)	0.214 (0.017)	396.15	406.63

Table 7. Simulation results based on bootstrap method for ZIPINGARCH(1,1) model

Method	Estimated parameter	Mean	Bias	SE	RMSE
QEF	$\hat{\alpha}_0$	0.289	0.012	0.007	0.014
	$\hat{\alpha}_1$	0.094	0.005	0.015	0.016
	$\hat{\beta}_1$	0.226	0.008	0.047	0.048
	$\hat{\omega}$	0.210	0.004	0.017	0.018
LEF	$\hat{\alpha}_0$	0.296	0.014	0.012	0.019
	$\hat{\alpha}_1$	0.109	0.025	0.024	0.034
	$\hat{\beta}_1$	0.245	0.012	0.067	0.068
	$\hat{\omega}$	0.198	0.010	0.018	0.021
ML	$\hat{\alpha}_0$	0.290	0.011	0.008	0.013
	$\hat{\alpha}_1$	0.095	0.008	0.015	0.016
	$\hat{\beta}_1$	0.237	0.024	0.049	0.054
	$\hat{\omega}$	0.199	0.015	0.018	0.023

Table 8. Diagnostics for ZIPINGARCH(1,1) model

	$LB_{30}(\epsilon_t)$	$LB_{30}(\epsilon_t^2)$
χ^2	25.4	24.22
p-value	0.705	0.762

This means that the ZIPINGARCH(1,1) model data. The cumulative periodogram plots further support the model’s accuracy (Figure 4).

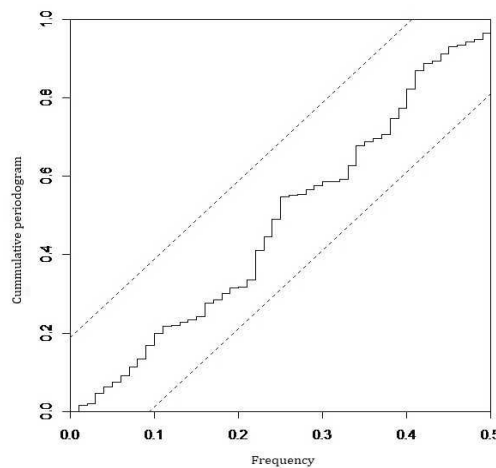


Fig. 4. Cummulative periodogram plot

5. Conclusions

This paper considers the QEF method for estimating the parameters of ZIP models. We have shown the superiority of the QEF method compared to the LEF method, theoretically.

Results also show that the information gain using the QEF method are more informative than that LEF method for the count data in ZIP models. Through the empirical studies, it is found that the ZIP models via QEF esti-

mates provide a better fit than the LEF and ML estimates. Hence, from the findings, the QEF method could serve as an alternative parameter estimation method in estimating the parameters for this class of count data models.

6. Acknowledgements

The authors wish to thank the editors and the referees for their valuable comments and suggestions. This research is financially supported by the Fundamental Research Grant Scheme (No. FP012-2013A).

References

- Abaza, B. (1982).** A recursive formulation for the two-step least squares estimator. *Kuwait Journal of Science*, **9**: 77-85.
- Allen, D., Ng, K.H. & Shelton, P. (2013).** Estimating and simulating Weibull models of risk or price durations: An application to ACD models. *North American Journal of Economics and Finance*, **25**: 214-225.
- Bahadur, R.R. (1958).** Examples of inconsistency of maximum likelihood estimates. *Sankhya: The Indian Journal of Statistics*, **20**: 207-210.
- Baksh, M.F., Bhning, D. & Lerdsuwansri, R. (2011).** An extension of an over-dispersion test for count data. *Computational Statistics and Data Analysis*, **55**(1): 466-474.
- Bera, A.K., Biliyas, Y. & Simlai, P. (2006).** Estimating functions and equations: An essay on historical developments with applications to econometrics. *Palgrave Handbook of Econometrics*, **1**: 427-476.
- Brockwell, P.J. & Davis, R.A. (1991).** *Time series: Theory and methods*. Springer Science & Business Media. New York. pp. 106-110.
- Chan, K.S. & Ledolter, J. (1995).** Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, **90**(429): 242-252.
- Chandra, S.A. & Taniguchi, M. (2001).** Estimating functions for nonlinear time series models. *Annals of the Institute of Statistical Mathematics*, **53**(1): 591-597.
- Crowder, M. (1987).** On linear and quadratic estimating functions. *Biometrika*, **74**(3): 591-597.
- Godambe, V.P. (1985).** The foundations of finite sample estimation in stochastic processes. *Biometrika*, **72**(2): 419-428.
- Godambe, V.P. & Heyde, C.C. (2010).** *Quasi-likelihood and optimal estimation. Selected Works of CC Heyde*, Springer. New York. pp. 386-399.
- Kharrati-Kopaei, M. & Faghih, H. (2011).** Inferences for the inflation parameter in the ZIP distributions: The method of moments. *Statistical Methodology*, **8**(4): 377-388.
- Lambert, D. (1992).** Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**(1): 1-14.
- Liang, Y., Thavaneswaran, A. & Abraham, B. (2011).** Joint estimation using quadratic estimating function. *Journal of Probability and Statistics*. DOI:10.1155/2011/372512.
- Merkouris, T. (2007).** Transform martingale estimating functions. *The Annals of Statistics*, **35**(5): 1975-2000.
- Nanjundan, G. & Naika, T.R. (2012).** Asymptotic comparison of method of moments estimators and maximum likelihood estimators of parameters in zero-inflated Poisson model. *Applied Mathematics*, **3**(6): 610-616.
- Ng, K.H. & Peiris, S. (2013).** Modelling high frequency transaction data in financial economics: A comparative study based on simulations. *Economic Computation and Economic Cybernetics Studies and Research*, **47**(2): 189-202.

- Ng, K.H., Shelton, P., Thavaneswaran, A. & Ng, K.H. (2015).** Modelling the risk or price durations in financial markets: Quadratic estimating functions and applications. *Economic Computation and Economic Cybernetics Studies and Research*, **49**(1): 223-238.
- Staub, K.E. & Winkelmann, R. (2012).** Consistent estimation of zero-inflated count models. *Health Economics*, **22**(6): 673-686.
- Thavaneswaran, A. & Abraham, B. (1988).** Estimation for nonlinear time series models using estimating equations. *Journal of Time Series Analysis*, **9**(1): 673-686.
- Thavaneswaran, A., Liang, Y. & Frank, J. (2012).** Inference for random coefficient volatility models. *Statistics and Probability Letters*, **82**(12): 2086-2090.
- Thavaneswaran, A., Ravishanker, N. & Liang, Y. (2015).** Generalized duration models and optimal estimation using estimating functions. *Annals of the Institute of Statistical Mathematics*, **67**(1): 129-156.
- Vinod, H. (1997).** Using Godambe-Durbin estimating functions in econometrics. *Lecture Notes-Monograph Series*. pp. 215-237. (<http://www.jstor.org/stable/4356018>)
- Zhu, F. (2012).** Zero-inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference*, **142**(4): 826-839.

Submitted: 02/11/2016

Revised : 08/02/2017

Accepted : 22/03/2017

التقدير الكفو في نماذج البواسون المتضخم عند الصفر (ZIP) وتطبيقاته في بيانات العد

¹نورول إن محمد، ^{2*}إبراهيم محمد، ³إن جي هاور، ⁴محمد يحيى
معهد العلوم الرياضية، كلية العلوم، جامعة مالايا، 50603 كوالا لامبور، ماليزيا
* imohamed@um.edu.my

الملخص

تم استخدام دوال التقدير لتقدير معلمات العديد من نماذج السلاسل الزمنية المستمرة. ومع ذلك، لم يتم تطبيق هذه المنهجية على نماذج بيانات العد. في هذا البحث، نستخدم في هذا البحث دوال التقدير من الدرجة الثانية (QEF) للحصول على تقدير مشترك لمعلمات المتوسط والتباين المشروطة لنماذج بيانات العد، وخاصة نموذج بواسون المتضخم عند الصفر (ZIP)، ونموذج ZIP للانحدار ونموذج الانحدار الذاتي غير المتجانس المعمم للأعداد الصحيحة التي يكون توزيعها الشرطي ZIP. تشير النتائج إلى أن التقديرات التي تم الحصول عليها باستخدام دوال التقدير من الدرجة الثانية، والتي تستخدم معلومات من دوال تقدير مزوجة، هي أكثر معلوماتية من دوال التقدير الخطية (LEF)، والتي تستخدم فقط معلومات من دوال التقدير. وأخيراً، قمنا بالتطبيق على بيانات حقيقية باستخدام نماذج ZIP وطرق QEF و LEF وطرق الاحتمال الأرجح، وتوضح هذه التطبيقات تقدم طريقة QEF على باقي الطرق عملياً.