Markov regression models for count time series with excess zeros: A partial likelihood approach

Article in Statistical Methodology · September 2013

DOI: 10.1016/j.stamet.2013.02.001

3 authors, including:

Gideon Zamba
University of Iowa

Joseph E. Cavanaugh
University of Iowa

Statistical Methodology 14 (2013) 26–38

Contents lists available at SciVerse ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Markov regression models for count time series with excess zeros: A partial likelihood approach

Ming Yang [a], Gideon K.D. Zamba [b], Joseph E. Cavanaugh [b,*]

[a] *Center for Biostatistics in AIDS Research, Harvard School of Public Health, United States*

[b] *Department of Biostatistics, College of Public Health, The University of Iowa, United States*

A R T I C L E I N F O

Article history:
Received 23 May 2012
Received in revised form
25 December 2012
Accepted 11 February 2013

Keywords:
Autoregressive
EM algorithm
Poisson distribution
Zero-inflation

A B S T R A C T

Count data with excess zeros are common in many biomedical and public health applications. The zero-inflated Poisson (ZIP) regression model has been widely used in practice to analyze such data. In this paper, we extend the classical ZIP regression framework to model count time series with excess zeros. A Markov regression model is presented and developed, and the partial likelihood is employed for statistical inference. Partial likelihood inference has been successfully applied in modeling time series where the conditional distribution of the response lies within the exponential family. Extending this approach to ZIP time series poses methodological and theoretical challenges, since the ZIP distribution is a mixture and therefore lies outside the exponential family. In the partial likelihood framework, we develop an EM algorithm to compute the maximum partial likelihood estimator (MPLE). We establish the asymptotic theory of the MPLE under mild regularity conditions and investigate its finite sample behavior in a simulation study. The performances of different partial-likelihood based model selection criteria are compared in the presence of model misspecification. Finally, we present an epidemiological application to illustrate the proposed methodology.

* Correspondence to: Department of Biostatistics, 105 River Street, N300 CPHB, The University of Iowa, Iowa City, IA, 52242, United States. Tel.: +1 319 384 1602; fax: +1 319 384 1591.

E-mail address: joe-cavanaugh@uiowa.edu (J.E. Cavanaugh).

1572-3127/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2013.02.001

1. Introduction

Time series data involving counts are frequently encountered in many biomedical and public health applications. For example, in disease surveillance, the occurrence of rare infections over time is often monitored by public health officials, and the time series collected can be used for the purpose of monitoring changes in disease activity. For rare diseases with low infection rates, the observed counts typically contain a high frequency of zeros (zero-inflation), but the counts can also be very large during an outbreak period. Failure to account for zero-inflation in the data may result in misleading inference and the detection of spurious associations.

Regression models based on the zero-inflated Poisson (ZIP) distribution have been well developed for independent count data [20,22]. To analyze repeated measures data with a large number of zeros, Hall [15] incorporated random effects in the classical ZIP regression model to account for within-subject correlation and between-subject heterogeneity. Marginal models [17] and multi-level models [21] have also been proposed to analyze clustered count data in the presence of zero inflation. To test whether the ZIP distribution should be used as an alternative to the ordinary Poisson distribution, a score test was first proposed for independent count data by Van den Broek [28], and was later extended to the setting where the count data are correlated [30]. A contemporary overview of zero-inflated models in the presence of overdispersion can be found in [24].

Despite the popularity of ZIP models, the literature for count time series with excess zeros is sparse. Yau et al. [31] presented a ZIP mixed autoregressive model and applied the model to evaluate a participatory ergonomics intervention in occupational health. Their model belongs to the class of parameter-driven models [5], where the time dependence between adjacent observations is characterized through an unobservable process. However, the first order autoregressive process employed for the latent process might be too restrictive to accurately approximate the actual temporal correlation in many series. Moreover, parameter estimation in parameter-driven models is often computationally burdensome [3,9], and it is thus difficult for such models to be routinely applied in practice. A comprehensive comparison of different estimation methods for parameter-driven models is presented by Nelson and Leroux [25].

In this paper, we propose an observation-driven model for zero-inflated count time series in which both the Poisson intensity and zero-inflation parameters can be time-varying. In the observation-driven model, time dependence is modeled as a function of past observations, and a general autoregressive structure can be easily accommodated. Note that observation-driven models are often referred to as transition or Markov models in the longitudinal data literature [8]. Observation-driven models for count series without zero-inflation have been previously investigated by Zeger and Qaqish [32], Davis et al. [6], Fokianos et al. [12], and Fokianos and Tjøstheim [13]. To model count time series with excess zeros, Wang [29] proposed a Markov ZIP model in which a two-state discrete time Markov chain is employed to accommodate the serial correlation. A recent paper by Zhu [33] extends the integer-valued generalized autoregressive conditional heteroscedasticity (GARCH) model to analyze zero-inflated and overdispersed count time series. In contrast to the approaches adopted by Wang [29] and Zhu [33], we propose a class of ZIP autoregressive models in a partial likelihood framework [4], which provides a unified and flexible approach to analyze non-Gaussian time series [19]. In the partial likelihood framework, both deterministic and stochastic covariate processes can be easily incorporated into the model. Partial likelihood inference has been successfully applied in modeling time series where the conditional distribution of the response lies within the exponential family. Extending this partial likelihood approach to ZIP time series poses methodological and theoretical challenges, since the ZIP distribution is a mixture and therefore lies outside the exponential family. Our proposed methodology is an extension of the work by Kedem and Fokianos [19] and Fokianos and Kedem [11] to a distribution outside the exponential family.

The paper is organized as follows. Section 2 introduces the ZIP autoregression under the partial likelihood framework. The numerical algorithm for parameter estimation is discussed in Section 3. In Section 4, we establish the large sample theory under mild regularity conditions; technical details are presented separately in the Appendix. Section 5 is devoted to variable selection in the presence of model misspecification. Section 6 features Monte Carlo simulations to evaluate the performance of parameter estimation and model selection in the finite sample situation. An application from infectious disease epidemiology is provided in Section 7 to illustrate the proposed methodology. Section 8 concludes the paper with a brief discussion of several future directions.

2. ZIP autoregression

In this section, we introduce an autoregressive model for count time series with excess zeros. The proposed model can be viewed as an extension of the Poisson autoregressive model discussed by Kedem and Fokianos [19, Chapter 4]. In the partial likelihood framework, we first derive the partial score process for the ZIP autoregression and then show that the process is a square integrable martingale. To obtain the standard errors of the estimators, the observed and conditional information matrices are also derived. The latter, defined by a cumulative sum of conditional covariance matrices, is akin to the well-known expected (Fisher) information matrix.

Let $\{Y_t\}_{t=1}^N$ denote the response series, composed of discrete count data. Suppose the count series is conditionally distributed as ZIP (λ_t, ω_t), with probability mass function (p.m.f.) defined as follows:

$$f_{Y_t}(y_t|\mathcal{F}_{t-1};\boldsymbol{\theta}) = \omega_t I_{(y_t=0)} + (1-\omega_t)\exp(-\lambda_t)\lambda_t^{y_t}/y_t!. \tag{2.1}$$

Here, $\mathcal{F}_{t-1}$ denotes a filtration that represents all that is known to the observer at time $t-1$ about the response series, and covariate information if available. The Poisson intensity and zero-inflation parameters are denoted by λ_t and ω_t, respectively. When $\omega_t = 0$, Eq. (2.1) reduces to the p.m.f. for the ordinary Poisson distribution. To simplify our notation, we use $y_{0,t}$ to represent $I_{(y_t=0)}$ for the rest of the paper.

For any non-negative integer k, the cumulative distribution function (c.d.f.) of $Y_t|\mathcal{F}_{t-1}$ is given by

$$F_{Y_t}(k|\mathcal{F}_{t-1};\boldsymbol{\theta}) = \sum_{y_t=0}^{k} f_{Y_t}(y_t|\mathcal{F}_{t-1};\boldsymbol{\theta}) = \omega_t + (1-\omega_t)\exp(-\lambda_t)\sum_{y_t=0}^{k}\lambda_t^{y_t}/y_t!.$$

Moreover the conditional mean and variance are given by

$$\mathrm{E}(Y_t|\mathcal{F}_{t-1};\boldsymbol{\theta}) = \lambda_t(1-\omega_t) \quad \text{and} \quad \mathrm{Var}(Y_t|\mathcal{F}_{t-1};\boldsymbol{\theta}) = \lambda_t(1-\omega_t)(1+\lambda_t\omega_t).$$

Since $\mathrm{Var}(Y_t|\mathcal{F}_{t-1};\boldsymbol{\theta})$ is greater than $\mathrm{E}(Y_t|\mathcal{F}_{t-1};\boldsymbol{\theta})$, the ZIP distribution can account for both zero-inflation and overdispersion in count data.

We propose a ZIP autoregressive model in which the parameters λ_t and ω_t are modeled as follows:

$$\log(\lambda_t) = \eta_t = \mathbf{x}_{t-1}^T\boldsymbol{\beta} \quad \text{and} \quad \log[\omega_t/(1-\omega_t)] = \xi_t = \mathbf{z}_{t-1}^T\boldsymbol{\gamma}, \tag{2.2}$$

where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^T$ and $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_q)^T$ are unknown parameters to be estimated. Here $\mathbf{x}_{t-1} = (x_{(t-1)1},\ldots,x_{(t-1)p})^T$ and $\mathbf{z}_{t-1} = (z_{(t-1)1},\ldots,z_{(t-1)q})^T$ denote vectors of past explanatory covariates, into which functions of the lagged response series can be incorporated to account for serial correlation. As with generalized linear models (GLMs), offset variables can be included in the above linear predictors whenever necessary.

Given a ZIP time series $\{Y_t\}_{t=1}^N$, the partial likelihood of the observed series is

$$\mathrm{PL}(\boldsymbol{\theta};\mathbf{y}) = \prod_{t=1}^{N} f_{Y_t}(y_t|\mathcal{F}_{t-1};\boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T,\boldsymbol{\gamma}^T)^T$ is the $p+q$ dimensional vector of unknown parameters. Unlike the full likelihood, the partial likelihood does not require complete knowledge of the joint distribution of the response and the covariates [19]. For the ZIP autoregression defined by Eqs. (2.1) and (2.2), we have

$$\log \mathrm{PL}(\boldsymbol{\theta};\mathbf{y}) = \sum_{t=1}^{N}\log[\omega_t y_{0,t} + (1-\omega_t)\exp(-\lambda_t)\lambda_t^{y_t}/y_t!].$$

The vector $\hat{\boldsymbol{\theta}}$ that maximizes the (log) partial likelihood is called the maximum partial likelihood estimator (MPLE). The partial score vector is then given by

$$\mathbf{S}_N(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log \mathrm{PL}(\boldsymbol{\theta}; \mathbf{y}) = \sum_{t=1}^{N} \mathbf{C}_{t-1}\mathbf{v}_t(\boldsymbol{\theta}),$$

with $\mathbf{C}_{t-1}$ and $\mathbf{v}_t(\boldsymbol{\theta})$ defined as follows:

$$\mathbf{C}_{t-1} = \begin{bmatrix} \mathbf{x}_{t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{t-1} \end{bmatrix} \quad \text{and} \quad \mathbf{v}_t(\boldsymbol{\theta}) = \begin{bmatrix} y_t - \lambda_t(1 - \omega_t y_{0,t}/p_{0,t}) \\ \omega_t(y_{0,t}/p_{0,t} - 1) \end{bmatrix}.$$

Here, $p_{0,t} = \omega_t + (1 - \omega_t)\exp(-\lambda_t)$ is the p.m.f. of $Y_t|\mathcal{F}_{t-1}$ at zero.

Solving the partial score equation $\mathbf{S}_N(\boldsymbol{\theta}) = \mathbf{0}$ for $\hat{\boldsymbol{\theta}}$ is a nontrivial task. In the next section, we will present an expectation maximization (EM) algorithm for parameter estimation that takes advantage of the finite mixture representation of the conditional ZIP distribution.

Since $\mathrm{E}(Y_t|\mathcal{F}_{t-1}) = \lambda_t(1 - \omega_t)$ and $\mathrm{E}(Y_{0,t}|\mathcal{F}_{t-1}) = p_{0,t}$, we have

$$\mathrm{E}[\mathbf{C}_{t-1}\mathbf{v}_t(\boldsymbol{\theta})|\mathcal{F}_{t-1}] = \mathbf{C}_{t-1}\mathrm{E}[\mathbf{v}_t(\boldsymbol{\theta})|\mathcal{F}_{t-1}] = \mathbf{0}.$$

Therefore, the partial score process $\{\mathbf{S}_t(\boldsymbol{\theta})\}$, defined by $\mathbf{S}_t(\boldsymbol{\theta}) = \sum_{s=1}^{t} \mathbf{C}_{s-1}\mathbf{v}_s(\boldsymbol{\theta})$, is a martingale satisfying

$$\mathrm{E}[\mathbf{S}_t(\boldsymbol{\theta})|\mathcal{F}_{t-1}] = \mathrm{E}[\mathbf{S}_{t-1}(\boldsymbol{\theta})|\mathcal{F}_{t-1}] + \mathrm{E}[\mathbf{C}_{t-1}\mathbf{v}_t(\boldsymbol{\theta})|\mathcal{F}_{t-1}] = \mathbf{S}_{t-1}(\boldsymbol{\theta}).$$

Moreover the martingale is square integrable such that

$$\mathrm{E}[\|\mathbf{S}_t(\boldsymbol{\theta})\|^2] \le \sum_{s=1}^{t} \mathrm{E}[\|\mathbf{C}_{s-1}\mathbf{v}_s(\boldsymbol{\theta})\|^2] = \sum_{s=1}^{t} \mathrm{E}\{\mathrm{E}[\|\mathbf{C}_{s-1}\mathbf{v}_s(\boldsymbol{\theta})\|^2|\mathcal{F}_{s-1}]\} < \infty.$$

These properties of the partial score process together with the martingale central limit theorem (CLT) ensure the consistency and asymptotic normality of the MPLE; further details are provided in Section 4.

Next we introduce the observed and conditional information matrices that can be used to compute standard errors of the MPLE. The observed information matrix (i.e., negative Hessian) of the ZIP autoregression is given by

$$\mathbf{H}_N(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log \mathrm{PL}(\boldsymbol{\theta}; \mathbf{y}) = \sum_{t=1}^{N} \mathbf{C}_{t-1}\mathbf{D}_t(\boldsymbol{\theta})\mathbf{C}_{t-1}^T,$$

where $\mathbf{D}_t(\boldsymbol{\theta})$ is a 2×2 matrix with elements defined as follows:

$$d_{11,t}(\boldsymbol{\theta}) = \lambda_t\{1 - y_{0,t}\omega_t[\omega_t + (1 - \omega_t)(1 + \lambda_t)\exp(-\lambda_t)]/p_{0,t}^2\},$$
$$d_{12,t}(\boldsymbol{\theta}) = -y_{0,t}\omega_t(1 - \omega_t)\lambda_t \exp(-\lambda_t)/p_{0,t}^2,$$
$$d_{22,t}(\boldsymbol{\theta}) = \omega_t(1 - \omega_t)[1 - y_{0,t}\exp(-\lambda_t)/p_{0,t}^2].$$

Similarly, the conditional information matrix (akin to the Fisher information) of the ZIP autoregression is given by:

$$\mathbf{G}_N(\boldsymbol{\theta}) = \sum_{t=1}^{N} \mathrm{Var}[\mathbf{C}_{t-1}\mathbf{v}_t(\boldsymbol{\theta})|\mathcal{F}_{t-1}] = \sum_{t=1}^{N} \mathbf{C}_{t-1}\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{C}_{t-1}^T,$$

where $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \mathrm{Var}[\mathbf{v}_t(\boldsymbol{\theta})|\mathcal{F}_{t-1}]$ is a 2×2 matrix with elements defined as follows:

$$\sigma_{11,t}(\boldsymbol{\theta}) = (1 - \omega_t)\lambda_t\{\exp(-\lambda_t) + \omega_t[1 - (1 + \lambda_t)\exp(-\lambda_t)]\}/p_{0,t},$$
$$\sigma_{12,t}(\boldsymbol{\theta}) = -\omega_t(1 - \omega_t)\lambda_t \exp(-\lambda_t)/p_{0,t},$$
$$\sigma_{22,t}(\boldsymbol{\theta}) = \omega_t^2(1 - \omega_t)[1 - \exp(-\lambda_t)]/p_{0,t}.$$

It is clear that both $\sigma_{11,t}(\boldsymbol{\theta})$ and $\sigma_{22,t}(\boldsymbol{\theta})$ are positive. Furthermore it can be verified that

$$\det\{\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\} = \sigma_{11,t}(\boldsymbol{\theta})\sigma_{22,t}(\boldsymbol{\theta}) - [\sigma_{12,t}(\boldsymbol{\theta})]^2 > 0.$$

Thus, $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is positive definite for all $\lambda_t \in (0, \infty)$ and $\omega_t \in (0, 1)$.

3. Parameter estimation

Maximizing the (log) partial likelihood to obtain the MPLE is a complicated nonlinear optimization problem. Although direct nonlinear optimization methods such as Newton–Raphson and Fisher scoring can be applied to locate the root of the partial score equation, such algorithms are often sensitive to starting values [14]. Moreover, they tend to either undershoot, overshoot, or cycle around the root in the presence of heavy zero inflation. To facilitate convergence, we take advantage of the finite mixture structure and estimate the unknown parameters through the EM algorithm [7], an iterative method widely used to fit statistical models involving latent variables.

For the ZIP autoregressive model introduced in Section 2, let u_t be a dichotomous variable indicating whether the observed count y_t comes from the degenerate distribution ($u_t = 1$) or the ordinary Poisson distribution ($u_t = 0$). The latent variable u_t is often unobservable and thus can be treated as missing in practice. Now consider the following equivalent representation of (2.1):

$$Y_t|u_t, \mathcal{F}_{t-1} \sim \text{Poisson}((1-u_t)\lambda_t)$$

and

$$u_t|\mathcal{F}_{t-1} \sim \text{Bernoulli}(\omega_t).$$

Applying Bayes theorem, we have

$$\mathrm{E}(u_t|y_t, \mathcal{F}_{t-1}; \lambda_t, \omega_t) = \mathrm{P}(u_t = 1|y_t, \mathcal{F}_{t-1}; \lambda_t, \omega_t) = \frac{\omega_t y_{0,t}}{\omega_t y_{0,t} + (1-\omega_t)\exp(-\lambda_t)\lambda_t^{y_t}/y_t!},$$

which constitutes the basis for the E-step.

The complete data log-partial likelihood can be orthogonally partitioned into

$$\log \mathrm{PL}^c(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u}) = \sum_{t=1}^{N} \log f(y_t|u_t, \mathcal{F}_{t-1}; \lambda_t) + \sum_{t=1}^{N} \log f(u_t|\mathcal{F}_{t-1}; \omega_t).$$

Since $[(1-u_t)\lambda_t]^{y_t} = y_{0,t}^{u_t}\lambda_t^{(1-u_t)y_t}$, then up to an additive constant,

$$\log \mathrm{PL}^c(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u}) = \sum_{t=1}^{N}(1-u_t)[y_t\log(\lambda_t) - \lambda_t] + \sum_{t=1}^{N}[u_t\log(\omega_t) + (1-u_t)\log(1-\omega_t)].$$

The EM algorithm seeks to find MPLE by iteratively applying the following two steps:

- E-step: Compute the expectation of $\log \mathrm{PL}^c(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u})$ with respect to the conditional distribution of $\mathbf{u}|\mathbf{y}, \boldsymbol{\theta}^{(j)}$. With $\hat{u}_t^{(j)} = \mathrm{E}(u_t|y_t, \mathcal{F}_{t-1}; \lambda_t^{(j)}, \omega_t^{(j)})$, we then have

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) = \sum_{t=1}^{N}(1-\hat{u}_t^{(j)})[y_t\log(\lambda_t) - \lambda_t] + \sum_{t=1}^{N}[\hat{u}_t^{(j)}\log(\omega_t) + (1-\hat{u}_t^{(j)})\log(1-\omega_t)].$$

- M-step: Find $\boldsymbol{\theta}^{(j+1)}$ that maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)})$. Due to the above orthogonal partition, we can obtain $\boldsymbol{\beta}^{(j+1)}$ and $\boldsymbol{\gamma}^{(j+1)}$ by maximizing

$$\sum_{t=1}^{N}(1-\hat{u}_t^{(j)})[y_t\log(\lambda_t) - \lambda_t]$$

and

$$\sum_{t=1}^{N}[\hat{u}_t^{(j)}\log(\omega_t) + (1-\hat{u}_t^{(j)})\log(1-\omega_t)]$$

separately. The M-step is equivalent to fitting two generalized linear models (i.e., Poisson and logistic regressions) and thus can be easily implemented in most existing statistical software.

The EM algorithm generally exhibits greater numerical stability than direct nonlinear optimization methods such as Newton–Raphson. In addition, wide overshoots or undershoots of the root are rare when the EM algorithm is used. The main drawback of the EM algorithm is that its convergence may be slow in certain applications (see [23], for more discussion). One can increase the speed of convergence by implementing an automatic switch to Newton–Raphson once the EM iterates begin to stabilize. As with any iterative optimization procedure, the EM algorithm may fail for starting values far removed from the point of convergence. Thus, despite the stability of the EM algorithm, one still needs to exercise care in choosing starting values.

4. Asymptotic results

In the previous sections, we introduced the ZIP autoregression under the partial likelihood framework, and the EM algorithm for the computation of the MPLE. In this section, we investigate the large sample behavior of the MPLE under the following regularity conditions (i.e., C.1–C.3). These conditions are slight modifications of those presented by Kedem and Fokianos [19]. They are assumed to hold throughout this section.

C.1 The true parameter $\boldsymbol{\theta}$ belongs to an open set $\Theta \subseteq R^{p+q}$.

C.2 The covariate matrix $\mathbf{C}_{t-1}$ almost surely lies in a non-random compact subset Γ of $R^{(p+q)\times 2}$ such that $\mathrm{P}(\sum_{t=1}^{N} \mathbf{C}_{t-1}\mathbf{C}_{t-1}^T$ is positive definite$) = 1$.

C.3 There is a probability measure ν on $R^{(p+q)\times 2}$ such that $\int \mathbf{C}\mathbf{C}^T \nu(d\mathbf{C})$ is positive definite, and such that for Borel sets $A \subset R^{(p+q)\times 2}$, at the true parameter $\boldsymbol{\theta}$,

$$\frac{1}{N}\sum_{t=1}^{N} I_{(\mathbf{C}_{t-1}\in A)} \xrightarrow{d} \nu(A).$$

Fokianos and Kedem [10] and Kedem and Fokianos [19, Chapter 3] provided a rigorous treatment of the asymptotic theory for non-stationary categorical time series. Their results are natural extensions of the work by Kaufmann [18] and their method of proof can be generally applied to any multivariate GLM. In Section 2, we have already verified that the partial score process for the ZIP autoregression is a square integrable martingale and that the conditional information matrix is positive definite. Therefore Lindeberg's and continuity conditions are satisfied, and the CLT for martingales [16] can be employed to prove the consistency and asymptotic normality of the MPLE.

Theorem 1. *Under the regularity conditions* C.1–C.3*, the MPLE for ZIP autoregression is consistent and*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}_{p+q}\left(\mathbf{0}, \mathbf{G}^{-1}(\boldsymbol{\theta})\right).$$

The proof of Theorem 1 parallels that of Kedem and Fokianos [19, pp. 130–134], and thus is not reproduced here for the sake of brevity. Applying the delta method to Theorem 1, we can obtain the asymptotic distributions of $\mathrm{E}(Y_t|\mathcal{F}_{t-1}; \hat{\boldsymbol{\theta}})$ and $F_{Y_t}(k|\mathcal{F}_{t-1}; \hat{\boldsymbol{\theta}})$ as stated in the following theorem.

Theorem 2. *Under the regularity conditions* C.1–C.3*, we have*

$$\sqrt{N}[\mathrm{E}(Y_t|\mathcal{F}_{t-1}; \hat{\boldsymbol{\theta}}) - \mathrm{E}(Y_t|\mathcal{F}_{t-1}; \boldsymbol{\theta})] \xrightarrow{d} \mathcal{N}_1\left(0, \mathbf{b}_t(\boldsymbol{\theta})^T \mathbf{C}_{t-1}^T \mathbf{G}^{-1}(\boldsymbol{\theta})\mathbf{C}_{t-1}\mathbf{b}_t(\boldsymbol{\theta})\right)$$

and

$$\sqrt{N}[F_{Y_t}(k|\mathcal{F}_{t-1}; \hat{\boldsymbol{\theta}}) - F_{Y_t}(k|\mathcal{F}_{t-1}; \boldsymbol{\theta})] \xrightarrow{d} \mathcal{N}_1\left(0, \mathbf{d}_t(\boldsymbol{\theta})^T \mathbf{C}_{t-1}^T \mathbf{G}^{-1}(\boldsymbol{\theta})\mathbf{C}_{t-1}\mathbf{d}_t(\boldsymbol{\theta})\right),$$

where $\mathbf{b}_t(\boldsymbol{\theta})$ *and* $\mathbf{d}_t(\boldsymbol{\theta})$ *are defined as follows:*

$$\mathbf{b}_t(\boldsymbol{\theta}) = \begin{bmatrix} \lambda_t(1-\omega_t) \\ -\lambda_t\omega_t(1-\omega_t) \end{bmatrix}, \qquad \mathbf{d}_t(\boldsymbol{\theta}) = \begin{bmatrix} (1-\omega_t)\exp(-\lambda_t)\displaystyle\sum_{y_t=0}^{k}(y_t-\lambda_t)\lambda_t^{y_t}/y_t! \\ \omega_t(1-\omega_t)\left[1-\exp(-\lambda_t)\displaystyle\sum_{y_t=0}^{k}\lambda_t^{y_t}/y_t!\right] \end{bmatrix}.$$

The proof of Theorem 2 is presented in the Appendix. The result of the theorem can be used to construct prediction intervals for $\mathrm{E}(Y_t|\mathcal{F}_{t-1};\boldsymbol{\theta})$ and $F_{Y_t}(k|\mathcal{F}_{t-1};\boldsymbol{\theta})$, given the information $\mathcal{F}_{t-1}$.

5. Model selection

Selecting an appropriate model among several competing candidates is a problem of great importance in many time series analyses. This task is often accomplished by using the Akaike [1] information criterion (AIC). AIC is derived as an estimator of the expected Kullback–Leibler discrepancy between the true model and a fitted candidate model. By replacing the traditional likelihood in the independent data setting by the partial likelihood, we have

$$\mathrm{AIC} = -2\log \mathrm{PL}(\hat{\boldsymbol{\theta}}; y) + 2\dim(\boldsymbol{\theta}),$$

where $\hat{\boldsymbol{\theta}}$ is the MPLE and $\dim(\boldsymbol{\theta})$ is the number of free parameters in the model. Despite its popularity, the asymptotic justification of AIC relies on a strong assumption that the true model is contained in the candidate class [2]. Unfortunately, this assumption is seldom satisfied since it is difficult, if possible, to have access to the generating model in most practical applications.

To relax the preceding assumption, Takeuchi [27] introduced the Takeuchi information criterion (TIC) as an attractive alternative to AIC. In the partial likelihood context, TIC is defined as

$$\mathrm{TIC} = -2\log \mathrm{PL}(\hat{\boldsymbol{\theta}}; y) + 2\mathrm{tr}\left[\mathbf{J}_N(\hat{\boldsymbol{\theta}})\mathbf{I}_N^{-1}(\hat{\boldsymbol{\theta}})\right],$$

where $\mathbf{I}_N^{-1}(\boldsymbol{\theta})$ is the covariance matrix of $\hat{\boldsymbol{\theta}}$ and $\mathbf{J}_N(\boldsymbol{\theta})$ is defined as follows:

$$\mathbf{J}_N(\boldsymbol{\theta}) = \sum_{t=1}^{N}\left[\frac{\partial \log f_{Y_t}(y_t|\mathcal{F}_{t-1};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\left[\frac{\partial \log f_{Y_t}(y_t|\mathcal{F}_{t-1};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]^T.$$

For the ZIP autoregressive model, we have

$$\mathbf{J}_N(\boldsymbol{\theta}) = \sum_{t=1}^{N}\mathbf{C}_{t-1}\mathbf{v}_t(\boldsymbol{\theta})\mathbf{v}_t(\boldsymbol{\theta})^T\mathbf{C}_{t-1}^T,$$

and the covariance matrix $\mathbf{I}_N^{-1}(\boldsymbol{\theta})$ can be replaced by either $\mathbf{H}_N^{-1}(\boldsymbol{\theta})$ or $\mathbf{G}_N^{-1}(\boldsymbol{\theta})$.

In general, the penalty term $2\mathrm{tr}\left[\mathbf{J}_N(\hat{\boldsymbol{\theta}})\mathbf{I}_N^{-1}(\hat{\boldsymbol{\theta}})\right]$ is very close to $2\dim(\boldsymbol{\theta})$ when the generating model is within the candidate collection. In such a setting, the performances of AIC and TIC are quite comparable. However, TIC often outperforms AIC in the presence of model misspecification, when none of the candidate models are correctly specified. When a candidate model is misspecified, the penalty term of TIC will often be much larger than that of AIC. This tendency prevents the selection of a fitted model that is too complicated for data at hand. It is worth noting that the form of TIC is very similar to the quasi-likelihood information criterion (QIC), which was developed by Pan [26] for generalized linear model selection where the GLMs are fit using generalized estimating equations (GEEs).

6. Simulation study

Two sets of simulation studies are featured in this section. We first investigate the finite sample behavior of the MPLE. We then compare the performances of AIC and TIC in the presence of model misspecification. For the remainder of this section, we assume the time series data is generated by the following model:

$$\eta_t = \beta_0 + \beta_1 I_{(y_{t-1}>0)} + \sigma z_t \quad \text{and} \quad \xi_t = \gamma_0 + \gamma_1 I_{(y_{t-1}>0)}. \tag{6.1}$$

Here, z_t is an unobservable realization from the standard normal distribution, included in η_t to introduce extra overdispersion into the data. Throughout the section, we assume $\boldsymbol{\theta} = (\beta_0, \beta_1, \gamma_0, \gamma_1)^T =$

Table 1
Finite sample results of the MPLE based on 1000 replications simulated independently from model (6.1) with $\sigma = 0$.

N	$\boldsymbol{\theta}$	Bias	ASE	ESD	CP
100	β_0	−0.012	0.133	0.135	0.958
	β_1	0.011	0.154	0.159	0.946
	γ_0	−0.016	0.303	0.297	0.961
	γ_1	0.028	0.426	0.417	0.959
200	β_0	−0.008	0.093	0.091	0.956
	β_1	0.007	0.108	0.107	0.956
	γ_0	−0.018	0.212	0.219	0.945
	γ_1	0.026	0.297	0.309	0.944
500	β_0	−0.004	0.058	0.059	0.957
	β_1	0.006	0.067	0.069	0.947
	γ_0	−0.011	0.133	0.130	0.956
	γ_1	0.022	0.187	0.179	0.963

$(1.2, 0.6, 0.4, -0.8)^T$ is the true parameter vector. In general, (6.1) can be viewed as a ZIP autoregressive model with unobservable random effects. In the special case when $\sigma = 0$, (6.1) reduces to the standard ZIP autoregressive model introduced in Section 2.

Table 1 summarizes the finite sample results of the MPLE under three different sample sizes ($N = 100, 200, 500$). From Table 1, we can see the absolute bias decreases as the sample size increases. The mean of the asymptotic standard error (ASE) and empirical standard deviation (ESD) are very close for all different sample sizes, and they both decrease as the sample size increases. In addition, the coverage probability (CP) of the 95% confidence interval is fairly close to the nominal level. Q–Q plots (not shown here) confirm the normality of the MPLE. It is worth noting the absolute bias, ASE, and ESD tend to be larger for the logistic component than for the log-linear component. This agrees with the existing literature, which indicates that it is advisable not to fit an overly complicated model for the zero-inflation parameter when sample information is limited or only a short sequence of observations is available [31].

We next investigate the variable selection problem. We assume the data is generated from model (6.1) with $\sigma = 0.5$. Our goal here is to compare the performances of AIC and TIC in the presence of model misspecification. To accomplish this, we consider the following nine candidate models:

$$\eta_t = \beta_0 + \sum_{i=1}^{k_1} \beta_i I_{(y_{t-i}>0)} \quad \text{and} \quad \xi_t = \gamma_0 + \sum_{i=1}^{k_2} \gamma_i I_{(y_{t-i}>0)}$$

for $k_1, k_2 = 0, 1, 2$. Here, all the candidate models are misspecified due to the omitted covariate z_t.

Table 2 summarizes the model selection results based on 1000 replications. Note that the true AR orders are $(k_1, k_2) = (1, 1)$. As the sample size increases from 100 to 500, the probability of jointly selecting the correct orders increases from 34.8% to 51.7% for AIC, and from 43.2% to 69.9% for TIC (Table 2).

7. Application

To illustrate the proposed methodology, we consider an application based on public health surveillance for syphilis, a sexually transmitted disease that remains a major public health challenge in the United States. According to the CDC, the rate of primary and secondary syphilis (the most infectious stages of the disease) decreased throughout the 1990s, and reached an all-time low in 2000. However, the syphilis rate has been increasing over the past decade, especially among men who have sex with men. For some states where syphilis is less common, the temporal surveillance for the disease is complicated due to the small counts collected over time.

Our data consists of the weekly number of syphilis cases reported in Maryland. It is extracted from the CDC's Morbidity and Mortality Weekly Report (MMWR). Fig. 1 displays the time series plot and histogram of the syphilis incidence for the period of 2007–2010, during which a large number of zeros

Table 2
Variable selection results of TIC and AIC (in parentheses) based on 1000 replications simulated independently from model (6.1) with $\sigma = 0.5$.

		$k_2 = 0$	$k_2 = 1$	$k_2 = 2$
$N = 100$	$k_1 = 0$	18 (11)	38 (12)	7 (4)
	$k_1 = 1$	220 (185)	**432** (**348**)	111 (95)
	$k_1 = 2$	57 (112)	91 (188)	26 (45)
$N = 200$	$k_1 = 0$	0 (0)	4 (2)	0 (0)
	$k_1 = 1$	66 (50)	**630** (**478**)	130 (92)
	$k_1 = 2$	13 (34)	126 (280)	31 (64)
$N = 500$	$k_1 = 0$	0 (0)	0 (0)	0 (0)
	$k_1 = 1$	0 (0)	**699** (**517**)	149 (106)
	$k_1 = 2$	0 (0)	127 (309)	25 (68)

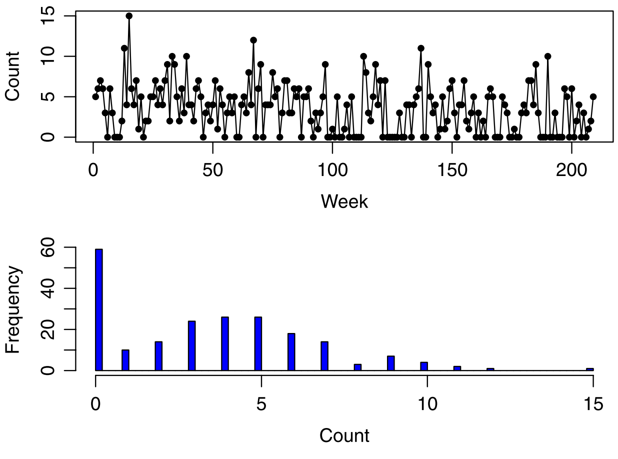

Fig. 1. Time series plot (top) and histogram (bottom) for weekly syphilis counts in Maryland from 2007 to 2010 (http://www.cdc.gov/mmwr/).

are observed over a total of 209 weeks. The time series plot seems to suggest a gradual decrease in the syphilis rate, but further investigation is needed to quantify the decrease and to formally test whether the downward trend is statistically significant.

To account for the serial autocorrelation and the preponderance of zeros, as evident by the peak at zero in the bimodal histogram, we consider the following ZIP autoregressive models:

$$\eta_t = \beta_0 + \sum_{i=1}^{k_1} \beta_i I_{(y_{t-i}>0)} + \beta_{k_1+1} x_t \quad \text{and} \quad \xi_t = \gamma_0 + \sum_{i=1}^{k_2} \gamma_i I_{(y_{t-i}>0)} + \gamma_{k_2+1} x_t, \tag{7.1}$$

for $k_1, k_2 = 0, \ldots, 4$. Here $x_t = t/1000$ represents the deterministic linear trend, which is always forced in the model since characterizing the trend is the primary objective of the study. For the case when $k_1 = k_2 = 0$, the model as described by (7.1) simply reduces to the classical ZIP regression without autocorrelation.

Fig. 2 displays the AIC and TIC values for all of the twenty-five candidate models. For each (k_1, k_2) combination, the TIC is always observed to be larger than the corresponding AIC. However, the most appropriate candidate model with $(k_1, k_2) = (1, 0)$ is favored by both AIC and TIC. Thus, our final model has the following structure:

$$\eta_t = \beta_0 + \beta_1 I_{(y_{t-1}>0)} + \beta_2 x_t \quad \text{and} \quad \xi_t = \gamma_0 + \gamma_1 x_t. \tag{7.2}$$

Table 3 summarizes the output for model (7.2) and its Poisson counterpart. Both models suggest an AR(1) component for the log-linear part, but the downward trend is found to be significant only

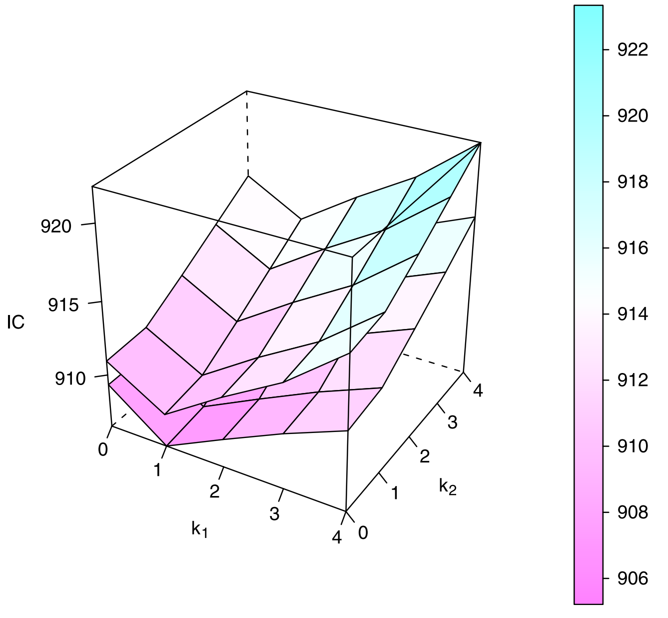

Fig. 2. AIC (bottom) and TIC (top) values for the 25 candidate models fit to the 2007–2010 syphilis data, with $k_1, k_2 = 0, \ldots, 4$. The penalties of TIC are uniformly larger than those of AIC. The model with $(k_1, k_2) = (1, 0)$ is favored by both AIC and TIC.

Table 3
Final ZIP autoregressive model and its Poisson counterpart for the Maryland syphilis data (2007–2010).

$\boldsymbol{\theta}$	ZIP model (TIC = 920.8, AIC = 918.8)			Poisson model (TIC = 1130.3, AIC = 1120.9)		
	Estimate	SE	*P*-value	Estimate	SE	*P*-value
β_0 (Intercept)	1.4894	0.1200	<0.0001	1.2822	0.1126	<0.0001
β_1 (AR1)	0.2211	0.1007	0.0281	0.3544	0.0952	0.0002
β_2 (Trend)	−1.0100	0.6669	0.1299	−3.1174	0.6448	<0.0001
γ_0 (Intercept)	−1.9332	0.3720	<0.0001			
γ_1 (Trend)	8.6052	2.8083	0.0022			

in the Poisson autoregressive model. When the zero-inflation part is included, we observe a huge reduction of AIC and TIC (more than 200), which indicates a pronounced lack-of-fit for the Poisson autoregressive model. Thus, the significant downward trend ($p < 0.0001$) obtained with this model should not be trusted. On the other hand, the ZIP autoregressive model also detects a downward but nonsignificant trend ($p = 0.1299$). According to this model, we observe more zeros in the latter period of series (Fig. 1) due to the significant increase of the zero-inflation parameter ($p = 0.0022$).

In practice, public health officials are often interested in forecasting future disease trends. This task can be easily accomplished based on (7.2). As an illustration, we consider one-step-ahead prediction where the model parameters are sequentially updated once a new observation becomes available. Since the number of syphilis cases in each week tends to be small, it is not very practical to use the predicted mean as a surveillance tool. As an alternative, we recommend using a predictive probability for the purpose of forecasting. Specifically, we sequentially compute the probability that the next count will be greater than a predetermined cutoff (e.g., 90% quantile). Such a forecast can be viewed as the predictive chance for a future outbreak. Fig. 3 features the plot of predictive probabilities (with a cutoff of 6) for the first 36 weeks of 2011 in Maryland. The ZIP autoregressive model closely tracks the average outbreak probability, while the Poisson autoregressive model, not surprisingly, tends to underestimate the average probability.

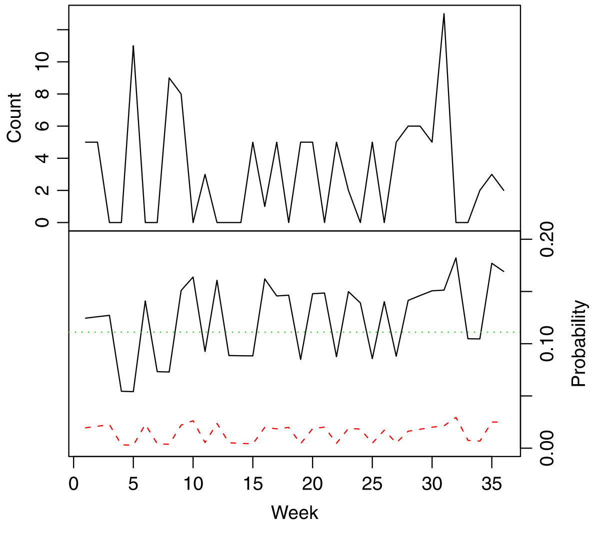

Fig. 3. Time series plot (top) and one-step-ahead predictive probabilities (bottom) for the first 36 weeks of 2011. The ZIP autoregressive model (black solid line) closely tracks the average outbreak probability (green dotted line), while the Poisson autoregressive model (red dashed line) tends to underestimate the average probability. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

We note that the forecasts for peaks in the series are generally off by a lag of one, a consequence of the autoregressive lag structure of the fitted model. According to the selection criteria, which are designed to choose an optimal predictive model, model (7.2) is the best forecasting model based on past values of the series. Additional complexity does not appear to improve prediction. Thus, it appears unlikely that a better forecasting model is available within the framework of models that are autoregressive in the structure for the intensity parameter and the zero-inflation probability. Prediction could potentially be improved by the incorporation of an appropriate covariate series that serves as a "leading indicator" of syphilis incidence. However, identifying such a series could prove to be challenging.

8. Conclusion

In this paper, we propose a ZIP autoregressive model to analyze count time series with excess zeros. The proposed model is a natural generalization of the Poisson autoregressive model discussed by Kedem and Fokianos [19, Chapter 4]. Partial likelihood is employed for statistical inference; thus, the fitted model should be interpreted conditionally. An EM algorithm has been developed for parameter estimation. An advantage of the EM algorithm is that it is robust to different starting values of the parameters. Large sample theory has been established to guarantee that the MPLE is consistent and asymptotically normal under mild regularity conditions. Simulation studies are reported to assess the finite sample properties of the MPLE, and to compare the performances of AIC and TIC as selection criteria in the presence of model misspecification. An application from disease surveillance is presented to illustrate the use of the ZIP autoregressive model. The proposed methodology can be extended to multivariate ZIP and zero-inflated negative binomial (ZINB) autoregressive models. In addition to the autoregressive components, it is also possible to include moving average components in the linear predictors to account for more complicated correlation structures in the data.

Acknowledgment

The authors wish to express their appreciation to an anonymous referee for valuable feedback which helped to improve the original version of this paper.

Appendix

In this Appendix we provide a proof for Theorem 2. At any time t, let

$$g_{1,t}(\boldsymbol{\theta}) = \mathrm{E}(Y_t|\mathcal{F}_{t-1};\boldsymbol{\theta}) = \lambda_t(1-\omega_t)$$

and

$$g_{2,t}(\boldsymbol{\theta};k) = F_{Y_t}(k|\mathcal{F}_{t-1};\boldsymbol{\theta}) = \omega_t + (1-\omega_t)\exp(-\lambda_t)\sum_{y_t=0}^{k}\lambda_t^{y_t}/y_t!.$$

We are interested in the large sample distributions of $g_{1,t}(\hat{\boldsymbol{\theta}})$ and $g_{2,t}(\hat{\boldsymbol{\theta}};k)$, where $\hat{\boldsymbol{\theta}}$ is the MPLE and it has been shown asymptotically normal in Theorem 1. Similar to the GLM setting, we may assert the following:

$$\frac{\partial\lambda_t}{\partial\eta_t} = \lambda_t, \qquad \frac{\partial\eta_t}{\partial\boldsymbol{\beta}} = \mathbf{x}_{t-1}, \qquad \frac{\partial\omega_t}{\partial\xi_t} = \omega_t(1-\omega_t), \quad \text{and} \quad \frac{\partial\xi_t}{\partial\boldsymbol{\gamma}} = \mathbf{z}_{t-1}.$$

The preceding will be repeatedly used in the subsequent derivations. Since $\partial g_{1,t}/\partial\lambda_t = 1-\omega_t$ and $\partial g_{1,t}/\partial\omega_t = -\lambda_t$, a direct application of the chain rule shows

$$\frac{\partial g_{1,t}}{\partial\boldsymbol{\beta}} = \frac{\partial g_{1,t}}{\partial\lambda_t}\frac{\partial\lambda_t}{\partial\eta_t}\frac{\partial\eta_t}{\partial\boldsymbol{\beta}} = [\lambda_t(1-\omega_t)]\,\mathbf{x}_{t-1} \tag{A.1}$$

and

$$\frac{\partial g_{1,t}}{\partial\boldsymbol{\gamma}} = \frac{\partial g_{1,t}}{\partial\omega_t}\frac{\partial\omega_t}{\partial\xi_t}\frac{\partial\xi_t}{\partial\boldsymbol{\gamma}} = [-\lambda_t\omega_t(1-\omega_t)]\,\mathbf{z}_{t-1}. \tag{A.2}$$

Moreover, it can be easily verified that

$$\frac{\partial g_{2,t}}{\partial\lambda_t} = (1-\omega_t)\exp(-\lambda_t)\sum_{y_t=0}^{k}(y_t/\lambda_t-1)\,\lambda_t^{y_t}/y_t!$$

and

$$\frac{\partial g_{2,t}}{\partial\omega_t} = 1-\exp(-\lambda_t)\sum_{y_t=0}^{k}\lambda_t^{y_t}/y_t!.$$

Thus, we have

$$\frac{\partial g_{2,t}}{\partial\boldsymbol{\beta}} = \frac{\partial g_{2,t}}{\partial\lambda_t}\frac{\partial\lambda_t}{\partial\eta_t}\frac{\partial\eta_t}{\partial\boldsymbol{\beta}} = \left[(1-\omega_t)\exp(-\lambda_t)\sum_{y_t=0}^{k}(y_t-\lambda_t)\lambda_t^{y_t}/y_t!\right]\mathbf{x}_{t-1} \tag{A.3}$$

and

$$\frac{\partial g_{2,t}}{\partial\boldsymbol{\gamma}} = \frac{\partial g_{2,t}}{\partial\omega_t}\frac{\partial\omega_t}{\partial\xi_t}\frac{\partial\xi_t}{\partial\boldsymbol{\gamma}} = \left\{\omega_t(1-\omega_t)\left[1-\exp(-\lambda_t)\sum_{y_t=0}^{k}\lambda_t^{y_t}/y_t!\right]\right\}\mathbf{z}_{t-1}. \tag{A.4}$$

Combining Eqs. (A.1)–(A.2) and (A.3)–(A.4) yields

$$\frac{\partial g_{1,t}}{\partial\boldsymbol{\theta}} = \begin{bmatrix}\mathbf{x}_{t-1} & \mathbf{0}\\ 0 & \mathbf{z}_{t-1}\end{bmatrix}\begin{bmatrix}\lambda_t(1-\omega_t)\\ -\lambda_t\omega_t(1-\omega_t)\end{bmatrix} = \mathbf{C}_{t-1}\mathbf{b}_t(\boldsymbol{\theta})$$

and

$$\frac{\partial g_{2,t}}{\partial\boldsymbol{\theta}} = \begin{bmatrix}\mathbf{x}_{t-1} & \mathbf{0}\\ 0 & \mathbf{z}_{t-1}\end{bmatrix}\begin{bmatrix}(1-\omega_t)\exp(-\lambda_t)\sum_{y_t=0}^{k}(y_t-\lambda_t)\,\lambda_t^{y_t}/y_t!\\ \omega_t(1-\omega_t)\left[1-\exp(-\lambda_t)\sum_{y_t=0}^{k}\lambda_t^{y_t}/y_t!\right]\end{bmatrix} = \mathbf{C}_{t-1}\mathbf{d}_t(\boldsymbol{\theta}).$$

Applying the delta method to Theorem 1, we have

$$\sqrt{N}[g_{1,t}(\hat{\boldsymbol{\theta}}) - g_{1,t}(\boldsymbol{\theta})] \xrightarrow{d} \mathcal{N}_1\left(0, \mathbf{b}_t(\boldsymbol{\theta})^T \mathbf{C}_{t-1}^T \mathbf{G}^{-1}(\boldsymbol{\theta}) \mathbf{C}_{t-1} \mathbf{b}_t(\boldsymbol{\theta})\right)$$

and

$$\sqrt{N}[g_{2,t}(\hat{\boldsymbol{\theta}}; k) - g_{2,t}(\boldsymbol{\theta}; k)] \xrightarrow{d} \mathcal{N}_1\left(0, \mathbf{d}_t(\boldsymbol{\theta})^T \mathbf{C}_{t-1}^T \mathbf{G}^{-1}(\boldsymbol{\theta}) \mathbf{C}_{t-1} \mathbf{d}_t(\boldsymbol{\theta})\right).$$

This completes the proof of Theorem 2.

References

[1] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Automatic Control 19 (1974) 716–723.
[2] J.E. Cavanaugh, Unifying the derivations of the Akaike and corrected Akaike information criteria, Statistics and Probability Letters 31 (1997) 201–208.
[3] K.S. Chan, J. Ledolter, Monte Carlo EM estimation for time series models involving counts, Journal of the American Statistical Association 90 (1995) 242–252.
[4] D.R. Cox, Partial likelihood, Biometrika 62 (1975) 269–276.
[5] D.R. Cox, Statistical analysis of time series: some recent developments, Scandinavian Journal of Statistics 8 (1981) 93–115.
[6] R.A. Davis, W.T.M. Dunsmuir, S.B. Streett, Observation-driven models for Poisson counts, Biometrika 90 (2003) 777–790.
[7] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood estimation from incomplete data via the EM algorithm, Journal of the Royal Statistical Society: Series B 39 (1977) 1–39.
[8] P.J. Diggle, P. Heagerty, K.-Y. Liang, S.L. Zeger, Analysis of Longitudinal Data, second ed., Oxford University Press, Oxford, 2002.
[9] J. Durbin, S.J. Koopman, Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives, Journal of the Royal Statistical Society: Series B 62 (2000) 3–56.
[10] K. Fokianos, B. Kedem, Prediction and classification of non-stationary categorical time series, Journal of Multivariate Analysis 67 (1998) 277–296.
[11] K. Fokianos, B. Kedem, Partial likelihood inference for time series following generalized linear models, Journal of Time Series Analysis 25 (2004) 173–197.
[12] K. Fokianos, A. Rahbek, D. Tjøstheim, Poisson autoregression, Journal of the American Statistical Association 104 (2009) 1430–1439.
[13] K. Fokianos, D. Tjøstheim, Log-linear Poisson autoregression, Journal of Multivariate Analysis 102 (2011) 563–578.
[14] G.H. Givens, J.A. Hoeting, Computational Statistics, Wiley, New Jersey, 2005.
[15] D.B. Hall, Zero-inflated Poisson and binomial regression with random effects: a case study, Biometrics 56 (2000) 1030–1039.
[16] P. Hall, C.C. Heyde, Martingale Limit Theory and its Applications, Academic Press, New York, 1980.
[17] D.B. Hall, Z. Zhang, Marginal models for zero inflated clustered data, Statistical Modelling 4 (2004) 161–180.
[18] H. Kaufmann, Regression models for nonstationary categorical time series: asymptotic estimation theory, The Annals of Statistics 15 (1987) 79–98.
[19] B. Kedem, K. Fokianos, Regression Models for Time Series Analysis, Wiley, New Jersey, 2002.
[20] D. Lambert, Zero-inflated Poisson regression, with an application to defects in manufacturing, Technometrics 34 (1992) 1–14.
[21] A.H. Lee, K. Wang, J.A. Scott, K.K.W. Yau, G.J. McLachlan, Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros, Statistical Methods in Medical Research 15 (2006) 47–61.
[22] A.H. Lee, K. Wang, K.K.W. Yau, Analysis of zero-inflated Poisson data incorporating extent of exposure, Biometrical Journal 43 (2001) 963–975.
[23] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, second ed., Wiley, New Jersey, 2007.
[24] J.M. Morel, N.K. Neerchal, Overdispersion Models in SAS, SAS Institute, North Carolina, 2012.
[25] K.P. Nelson, B.G. Leroux, Statistical models for autocorrelated count data, Statistics in Medicine 25 (2006) 1413–1430.
[26] W. Pan, Akaike's information criterion in generalized estimating equations, Biometrics 57 (2001) 120–125.
[27] K. Takeuchi, Distribution of information statistics and criteria for adequacy of models, Mathematical Science 153 (1976) 12–18.
[28] J. Van den Broek, A score test for zero inflation in a Poisson distribution, Biometrics 51 (1995) 738–743.
[29] P. Wang, Markov zero-inflated Poisson regression models for a time series of counts with excess zeros, Journal of Applied Statistics 28 (2001) 623–632.
[30] L. Xiang, A.H. Lee, K.K.W. Yau, G.J. McLachlan, A score test for zero-inflation in correlated count data, Statistics in Medicine 25 (2006) 1660–1671.
[31] K.K.W. Yau, A.H. Lee, P.J.W. Carrivick, Modeling zero-inflated count series with application to occupational health, Computer Methods and Programs in Biomedicine 74 (2004) 47–52.
[32] S.L. Zeger, B. Qaqish, Markov regression models for time series: a quasi-likelihood approach, Biometrics 44 (1988) 1019–1031.
[33] F. Zhu, Zero-inflated Poisson and negative binomial integer-valued GARCH models, Journal of Statistical Planning and Inference 142 (2012) 826–839.