We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 9

State-Space Models for Binomial Time Series with Excess Zeros

Fan Tang and Joseph E. Cavanaugh

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71336

Abstract

Count time series with excess zeros are frequently encountered in practice. In characterizing a time series of counts with excess zeros, two types of models are commonplace: models that assume a Poisson mixture distribution, and models that assume a binomial mixture distribution. Extensive work has been published dealing with modeling frameworks based on Poisson-type approaches, yet little has concentrated on binomial-type methods. To handle such data, we propose two general classes of time series models: a class of observation-driven ZIB (ODZIB) models, and a class of parameter-driven ZIB (PDZIB) models. The ODZIB model is formulated in the partial likelihood framework, which facilitates model fitting using standard statistical software for ZIB regression models. The PDZIB model is conveniently formulated in the state-space framework. For parameter estimation, we devise a Monte Carlo Expectation Maximization (MCEM) algorithm, with particle filtering and particle smoothing methods employed to approximate the intractable conditional expectations in the E-step of the algorithm. We investigate the efficacy of the proposed methodology in a simulation study, which compares the performance of the proposed ZIB models to their counterpart zero-inflated Poisson (ZIP) models in characterizing zero-inflated count time series. We also present a practical application pertaining to disease coding.

Keywords: autocorrelation, count time series, observation-driven models, parameter-driven-models, particle methods, zero-inflation

1. Introduction

Count time series with excess zeros are commonly encountered in a variety of research fields. In principle, both zero-inflation and autocorrelation may be present in such series. Failing to adequately accommodate temporal dynamics and a high frequency of zeros can lead to

INTECH
open science | open minds

© 2018 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

incorrect inferential conclusions. Developing a general modeling framework that accounts for these characteristics poses a daunting challenge.

In characterizing data comprised of counts with excess zeroes, two types of models are commonplace: a model that assumes a Poisson mixture distribution, and a model that assumes a binomial mixture distribution. A considerable literature exists for regression models based on the zero-inflated Poisson (ZIP) distribution to deal with count data that are independently distributed [1]. Many researchers have extended the classical ZIP model to analyze repeated measures data by incorporating independent random effects, as these can account for within-subject correlation and between-subject heterogeneity [2, 3]. To deal with count time series with excess zeros, some researchers have proposed parameter-driven ZIP models that accommodate the temporal dynamics by incorporating correlated random effects, which can be represented by a latent autoregressive process [4, 5]. However, for data arising from a binomial mixture distribution, a survey of the literature for analogous frameworks reflects an absence of work dealing with binomial time series with excess zeros. To handle such data, we propose two general classes of models: a class of observation-driven ZIB (ODZIB) models, and a class of parameter-driven ZIB (PDZIB) models. The inspiration for the two proposed modeling frameworks arises from the work of Hall [6], Yau et al. [4], and Yang et al. [5, 7].

Depending on how the temporal correlation is conceptualized, count time series models can be classified as either observation-driven or parameter-driven [8]. For the former, serial correlation is characterized by specifying that the conditional mean of the current response depends explicitly on its past values [9–14]. For the latter, such correlation is characterized through an unobservable underlying process [15–19]. In this chapter, we employ the partial likelihood framework to formulate the ODZIB model, as this largely simplifies parameter estimation with negligible loss of information. The ODZIB model can be viewed as an extension of the observation-driven binomial model [20]. Such a model is often fit using standard statistical software available for classical ZIB regression models. For the PDZIB model, we employ a state-space approach, as this framework allows for the investigation of the underlying latent processes that govern the temporal correlation and zero inflation. Due to the non-Gaussian distribution of the count response, and the nonlinear nature of modeling its conditional mean, traditional state-space methods using the Kalman filter and the Kalman smoother are not available for parameter estimation. We thereby adopt a Monte Carlo Expectation Maximization (MCEM) algorithm based on the particle filter [21] and the particle smoother [22].

The remainder of the chapter is organized as follows. In Section 2, we briefly introduce a class of observation-driven models for a zero-inflated count time series that arises from a binomial mixture. Section 3 proposes a class of parameter-driven models in the state-space framework, and presents the MCEM algorithm devised to fit such models. A comprehensive simulation study is provided in Section 4. In Section 5, we illustrate the proposed methodology through a practical application. Section 6 concludes with a brief discussion.

2. Observation-driven ZIB models

2.1. ZIB models

A popular approach for modeling independent zero-inflated binomial data is the ZIB model proposed by Hall [6]. This model assumes that data are generated from a mixture distribution, comprised of a binomial distribution and a degenerate distribution at zero. For response variable Y, let y_i denote the observation for subject i, $i = 1, 2, \ldots, n$. The probability mass function for the ZIB model is defined as follows:

$$f(y_i|\pi_i, \omega_i) = \begin{cases} \omega_i + (1-\omega_i)(1-\pi_i)^{n_i}, & \text{if } y_i = 0, \\ (1-\omega_i)\binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}, & \text{if } y_i > 0. \end{cases} \tag{1}$$

Here, ω_i is the zero-inflation parameter, and π_i is the intensity parameter representing the probability of success, both modeled via logit link functions:

$$\text{logit}(\omega_i) = \mathbf{x}_{i1}^T\gamma, \tag{2}$$

$$\text{logit}(\pi_i) = \mathbf{x}_{i2}^T\beta. \tag{3}$$

In the preceding, $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$ are sets of explanatory variables for the corresponding vectors of regression coefficients γ and β. The Expectation Maximization (EM) algorithm or the Newton-Raphson method can be used to obtain the parameter estimates.

2.2. Observation-driven ZIB models

In this section, we introduce an autoregressive model for binomial time series with excess zeros based on an observation-driven approach. We retain the same model structure as that introduced in Section 2.1 to account for the binomial mixture, yet we employ lagged responses as covariates to resolve the temporal correlation. The proposed model can be viewed as an extension of the binomial time series model presented by Kedem and Fokianos [20].

Let y_t denote the binomial count response. Define the information set

$$\mathcal{F}_{t-1} = \sigma\{y_{t-1}, y_{t-2}, \ldots, \mathbf{x}_t\} \tag{4}$$

so as to represent all that is known to the observer at time t about the response and any relevant covariate processes. Thus, the vector $\mathbf{x}_t$ represents a collection of past and possibly present time-dependent covariates that are observed at time $t-1$. In the present setting, $\mathbf{x}_t$ may be viewed as either fixed or random. Conditioning on the information $\mathcal{F}_{t-1}$, the response is assumed to follow a ZIB distribution with probability mass function defined as follows:

$$f_t(y_t|\mathcal{F}_{t-1};\pi_t,\omega_t) = \begin{cases} \omega_t + (1-\omega_t)(1-\pi_t)^{n_t}, & \text{if } y_t = 0, \\ (1-\omega_t)\binom{n_t}{y_t}\pi_t^{y_t}(1-\pi_t)^{n_t-y_t}, & \text{if } y_t > 0. \end{cases} \tag{5}$$

Similarly, ω_t and π_t represent the zero-inflation parameter and the intensity parameter, respectively. Both parameters are modeled via logit link functions. Specifically, we assume that

$$\text{logit}(\omega_t) = \mathbf{x}_{1,t}^T\gamma, \tag{6}$$

$$\text{logit}(\pi_t) = \mathbf{x}_{2,t}^T\beta + \sum_{j=1}^{p}\phi_j y_{t-j}, \tag{7}$$

where $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$ are sets of time-dependent explanatory variables for the corresponding vectors of regression coefficients γ and β, and $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_p]^T$ is a vector of autoregressive coefficients corresponding to the past responses $[y_{t-1}, \ldots, y_{t-p}]^T$. For simplicity, we treat the zero-inflation parameter ω_t as a constant that does not vary over time. In the observation-driven ZIB model, serial correlation is accommodated by introducing lagged values of the response to the linear predictor.

The partial data likelihood of the observed series is

$$\text{PL}(\boldsymbol{\theta}) = \prod_{t=1}^{n} f_t(y_t|\mathcal{F}_{t-1}), \tag{8}$$

where $\boldsymbol{\theta} = [\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\gamma}]^T$ is the vector of unknown parameters. The partial likelihood does not require the derivation of the joint distribution of the response and the covariates, and is largely simplified relative to the full likelihood. This approach facilitates conditional inference for a fairly large class of transitional processes where the response depends on its past values.

The log-likelihood for the observation-driven ZIB model is

$$\log \text{PL}(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log\left\{\omega_t I_{(y_t=0)} + (1-\omega_t)\binom{n_t}{y_t}\pi_t^{y_t}(1-\pi_t)^{n_t-y_t}\right\}. \tag{9}$$

The vector $\hat{\theta}$ obtained by maximizing the partial likelihood is called the maximum partial likelihood estimator (MPLE).

Similar to Section 2.1, we can apply the EM algorithm or the Newton–Raphson method to obtain the MPLE. This estimation process can be conveniently conducted in practice using standard software tools available for fitting classical ZIB models. In SAS, we can use the finite mixture models (FMM) procedure to fit the observation-driven ZIB model, while we can use function gamlss in the package generalized additive models for location scale and shape (GAMLSS) for model fitting in R. Hypothesis testing for $\boldsymbol{\theta}$ is carried out through the partial likelihood method. The common tests are based on Wald statistics, score statistics, and partial likelihood ratio statistics. All of these tests are conducted based on the framework for classical maximum likelihood inference.

3. Parameter-driven ZIB models

3.1. Model formulation

An alternative approach to describe binomial time series with excess zeros is based on parameter-driven ZIB models. This class of models can be viewed as an analogue of the parameter-driven ZIP models presented by Yang et al. [5].

To account for temporal dynamics in the series, we introduce a latent stationary autoregressive process $\{z_t\}$ of order p (AR(p)):

$$z_t = \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t. \tag{10}$$

Here, ε_t is a Gaussian white noise process with a mean of 0 and a variance of σ^2. Additionally, ϕ_i explains how the past state z_{t-i} relates to the current state z_t.

Let y_t be the observed count at time t. Given the current state z_t, the positive count response y_t is assumed to follow a ZIB distribution with a probability mass function defined as

$$f_t(y_t|z_t;\pi_t,\omega_t) = \begin{cases} \omega_t + (1-\omega_t)(1-\pi_t)^{n_t}, & \text{if } y_t = 0, \\ (1-\omega_t)\binom{n_t}{y_t}\pi_t^{y_t}(1-\pi_t)^{n_t-y_t}, & \text{if } y_t > 0. \end{cases} \tag{11}$$

Similar to the previous model parameterizations, ω_t and π_t represent the zero-inflation parameter and the intensity parameter, respectively. Both parameters are modeled via logit link functions and could be time-varying. To relate the intensity parameter π_t to the latent component z_t, we use the model

$$\text{logit}(\pi_t) = \mathbf{x}_t^{\mathrm{T}}\beta + z_t, \tag{12}$$

where $\mathbf{x}_t$ is a set of explanatory variables observed at time t, and β is the corresponding vector of regression coefficients. In the present setting, $\mathbf{x}_t$ is assumed fixed. For simplicity, we treat the zero-inflation parameter ω_t as a constant that does not vary over time.

For the parameter-driven ZIB model, the conditional mean and variance of the response variable y_t are given by

$$\mathrm{E}(Y_t|z_t) = (1\text{-}\omega_t)n_t\pi_t, \tag{13}$$

$$\mathrm{Var}(Y_t|z_t) = (1\text{-}\omega_t)n_t\pi_t[1\text{-}\pi_t(1\text{-}\omega_t n_t)]. \tag{14}$$

Obviously, the presence of zero-inflation ($\omega_t > 0$) not only explains the excess zeros in the series, but also introduces overdispersion. Additionally, the correlated random effects z_t contribute to the extra variance.

We can write the parameter-driven ZIB model in the following hierarchical form:

$$\mathbf{s}_t|\mathbf{s}_{t-1} \sim \mathcal{N}_p(\mathbf{\Phi}\mathbf{s}_{t-1}, \mathbf{\Sigma}), \tag{15}$$

$$u_t \sim \text{Bernoulli}\,(\omega), \tag{16}$$

$$y_t|\mathbf{s}_t, u_t \sim \text{Binomial}\,(n_t, (1-u_t)\pi_t), \tag{17}$$

where $\mathbf{s}_t = [z_t, \ldots, z_{t-p+1}]^T$ is a p-dimensional state vector with z_t being its first element, u_t is an unobservable membership indicator that determines whether the response comes from a degenerate distribution or an ordinary binomial distribution, $\boldsymbol{\Phi}$ is an unknown transition matrix, and $\boldsymbol{\Sigma}$ is the covariance matrix of the state noise process $\mathbf{s}_t$. The process $\mathbf{s}_t$ is initiated with a normal vector $\mathbf{s}_0$ that has mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. Diffuse priors are often assigned to $\mathbf{s}_0$ in practice. Given the two unobserved latent processes $\mathbf{s}_t$ and u_t, we can conceptualize a sequential update of the response variable y_t.

In Eq. (15), $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ are $p \times p$ matrices defined as follows:

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{18}$$

The transition matrix $\boldsymbol{\Phi}$ governs the generation of the state vector $\mathbf{s}_t$ from the past state $\mathbf{s}_{t-1}$ for time points $t = 1, \ldots, n$. Note that the covariance matrix $\boldsymbol{\Sigma}$ in Eq. (18) is not positive definite. This is both legitimate and common in the state-space modeling approach.

3.2. Parameter estimation via MCEM algorithm

3.2.1. Model fitting

To fit the parameter-driven ZIB model, in principle, one would first obtain the marginal likelihood of the observed data $y_1, \ldots, y_n$ by integrating out unobserved components. However, because of the presence of correlated random effects and the non-Gaussian nature of the response, these integrals are not analytically tractable. Therefore, approximations or numerical solutions for the maximum likelihood estimates (MLEs) are necessary. Instead of obtaining the MLEs based on the marginal likelihood, we propose an EM algorithm [23], which relies on the complete-data likelihood to estimate the parameters.

Let $y_{1:t} = [y_1, y_2, \ldots, y_t]^T$ denote the vector of observed data from time point 1 through t. In a similar fashion, let $\mathbf{s}_{0:t} = [\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_t]^T$ and $u_{1:t} = [u_1, u_2, \ldots, u_t]^T$ denote the vectors of two latent processes, respectively, over the same time frame. Let $\boldsymbol{\theta} = [\omega, \beta^T, \phi^T, \sigma]^T$ denote the vector of unknown parameters.

To develop an EM algorithm for parameter estimation of the mixture model, Eqs. (15)–(17), we begin by formulating the complete-data likelihood; i.e., the joint density of $\mathbf{s}_{0:n}$, $u_{1:n}$, and $y_{1:n}$. The two latent processes $\mathbf{s}_{0:n}$ and $u_{1:n}$ are considered missing. Based on the state-space representation, the complete-data likelihood may be orthogonally decomposed as follows:

$$\begin{aligned} L_c(\boldsymbol{\theta}) &= f\left(\mathbf{s}_{0:n}, u_{1:n}, y_{1:n}\right) \\ &= f(\mathbf{s}_{0:n}, u_{1:n}) f\left(y_{1:n} | \mathbf{s}_{0:n}, u_{1:n}\right) \\ &= f(\mathbf{s}_{0:n}) f(u_{1:n}) f\left(y_{1:n} | \mathbf{s}_{0:n}, u_{1:n}\right) \\ &= f(\mathbf{s}_0) \prod_{t=1}^{n} f(\mathbf{s}_t | \mathbf{s}_{t-1}) \prod_{t=1}^{n} f(u_t) \prod_{t=1}^{n} f\left(y_t | \mathbf{s}_t, u_t\right). \end{aligned} \tag{19}$$

Here, the initial state vector $\mathbf{s}_0$ is assumed to be normally distributed with mean vector μ_0 and covariance matrix $\boldsymbol{\Sigma}_0$. In implementing the algorithm, we set $\boldsymbol{\mu}_0 = 0$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_p$, as the effect of the starting values of $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ on the estimated parameters $\boldsymbol{\theta}$ is negligible.

Up to an additive constant, the complete-data log-likelihood is given by

$$\begin{aligned} l_c(\boldsymbol{\theta}) = &-\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{n} \left(z_t - \phi^T \mathbf{s}_{t-1}\right)^2 \\ &+ \sum_{t=1}^{n} \{u_t \log \omega + (1 - u_t) \log (1 - \omega)\} \\ &+ \sum_{t=1}^{n} (1 - u_t) \left\{y_t \mathbf{x}_t^T \beta - n_t \log \left(1 + \exp\left(\mathbf{x}_t^T \beta + z_t\right)\right)\right\}. \end{aligned} \tag{20}$$

The complete-data log-likelihood can be described as the sum of three functionally independent parameter forms, such that $l_c(\boldsymbol{\theta}) = l(\phi, \sigma \mid \mathbf{s}_t) + l(\omega \mid u_t) + l(\beta \mid \mathbf{s}_t, u_t)$, resulting in ease of the maximization in the M-step for each set of parameters.

With the implementation of the EM algorithm, we need to compute the conditional expectation of $l_c(\boldsymbol{\theta})$ given the observed data $y_{1:n}$. Deriving an analytical form for the conditional expectation is not feasible due to the nonlinear forms in the latent variables and the response, as well as the non-Gaussian distributions of the response and the latent indicators. There are many numerical methods available to approximate the conditional expectation, such as the Markov chain Monte Carlo (MCMC) algorithm [24], the MCEM algorithm [7, 25], the penalized quasi-likelihood (PQL) method [2], and integrated nested Laplace approximations (INLA) [26]. Following Yang et al. [5], we develop an MCEM algorithm to approximate the conditional expectation.

To simplify the notation, we let $\mathbf{A}_t^{(j)}$, $\mathbf{b}_t^{(j)}$, $c_t^{(j)}$, $d_t^{(j)}$, $e_t^{(j)}$, and $f_t^{(j)}$ denote the conditional expectations of $\mathbf{s}_{t-1}\mathbf{s}_{t-1}^T$, $z_t \mathbf{s}_{t-1}$, z_t^2, u_t, $(1 - u_t) \log \left(1 + \exp\left(\mathbf{x}_t^T \beta + z_t\right)\right)$, and $(1 - u_t) \exp\left(\mathbf{x}_t^T \beta + z_t\right) / \left(1 + \exp\left(\mathbf{x}_t^T \beta + z_t\right)\right)$ evaluated at $\boldsymbol{\theta}^{(j)}$, respectively. In the Monte Carlo E-step of the algorithm, we first compute the conditional expectation of $l_c(\boldsymbol{\theta})$:

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}\right) &= E\left\{l_c(\boldsymbol{\theta})|y_{1:n}, \boldsymbol{\theta}^{(j)}\right\} \\ &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}\left(c_t^{(j)} - 2\phi^T\mathbf{b}_t^{(j)} + \phi^T\mathbf{A}_t^{(j)}\phi\right) \\ &+ \sum_{t=1}^{n}\left\{d_t^{(j)}\log\omega + \left(1-d_t^{(j)}\right)\log(1-\omega)\right\} \\ &+ \sum_{t=1}^{n}\left\{\left(1-d_t^{(j)}\right)y_t\mathbf{x}_t^T\beta - n_t e_t^{(j)}\right\}, \end{aligned} \tag{21}$$

where particle filtering and smoothing techniques are used to approximate the conditional expectations. The details of the particle methods for the parameter-driven ZIB model are presented in Section 3.3.

The following partial derivatives are applied to maximize $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)})$ in the M-step:

$$\frac{\partial Q}{\partial\omega} = \frac{1}{\omega}\sum_{t=1}^{n} d_t^{(j)} - \frac{1}{1-\omega}\sum_{t=1}^{n}\left(1-d_t^{(j)}\right), \tag{22}$$

$$\frac{\partial Q}{\partial\phi} = \frac{1}{\sigma^2}\sum_{t=1}^{n}\left(\mathbf{b}_t^{(j)} - \mathbf{A}_t^{(j)}\phi\right), \tag{23}$$

$$\frac{\partial Q}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{t=1}^{n}\left(c_t^{(j)} - 2\phi^T\mathbf{b}_t^{(j)} + \phi^T\mathbf{A}_t^{(j)}\phi\right), \tag{24}$$

$$\begin{aligned} \frac{\partial Q}{\partial\beta} &= \frac{\partial E\left\{l_c(\boldsymbol{\theta})|y_{1:n}, \boldsymbol{\theta}^{(j)}\right\}}{\partial\beta} \\ &= E\left(\frac{\partial l_c(\boldsymbol{\theta})}{\partial\beta}|y_{1:n}, \boldsymbol{\theta}^{(j)}\right) \\ &= E\left(\sum_{t=1}^{n}\left\{(1-u_t)y_t - n_t(1-u_t)\frac{\exp\left(\mathbf{x}_t^T\beta + z_t\right)}{1+\exp\left(\mathbf{x}_t^T\beta + z_t\right)}\right\}\mathbf{x}_t|y_{1:n}, \boldsymbol{\theta}^{(j)}\right) \\ &= \sum_{t=1}^{n}\left\{\left(1-d_t^{(j)}\right)y_t - n_t f_t^{(j)}\right\}\mathbf{x}_t. \end{aligned} \tag{25}$$

At the j^{th} iteration, we obtain the following closed-form solutions for $\omega^{(j+1)}$, $\phi^{(j+1)}$, and $\sigma^{(j+1)}$:

$$\omega^{(j+1)} = \frac{1}{n}\sum_{t=1}^{n} d_t^{(j)}, \tag{26}$$

$$\phi^{(j+1)} = \left(\sum_{t=1}^{n}\mathbf{A}_t^{(j)}\right)^{-1}\sum_{t=1}^{n}\mathbf{b}_t^{(j)}, \tag{27}$$

$$\sigma^{(j+1)} = \sqrt{\frac{1}{n}\left\{\sum_{t=1}^{n} a_t^{(j)} - \left(\sum_{t=1}^{n} \mathbf{b}_t^{(j)}\right)^T \left(\sum_{t=1}^{n} \mathbf{A}_t^{(j)}\right)^{-1} \sum_{t=1}^{n} \mathbf{b}_t^{(j)}\right\}}. \tag{28}$$

In addition, we can easily compute $\beta^{(j+1)}$ through iterative algorithms such as Broyden-Fletcher-Goldfarb-Shanno (BFGS). Once we acquire the particle smoothers from the smoothing step, we can obtain the MCEM estimates by plugging in the sample means of the functions of particle smoothers for the conditional expectations.

To offset the slow convergence and to reduce the computational cost of the EM algorithm, starting with good initial parameters is essential. For the proposed parameter-driven ZIB model, we suggest using the estimates of the parameters from a classical ZIB model or from the observation-driven ZIB model discussed in Section 2.2.

3.2.2. Standard errors

Standard errors of the parameter estimators can be obtained either by using the inverse of the observed information to approximate the variance/covariance matrix, or by employing a collection of replicated bootstrapped parameter estimates. Given the computational cost of the MCEM algorithm, we pursue the first approach by applying Louis's formula [27] to compute the observed information matrix $\mathbf{I}_o(\boldsymbol{\theta})$. Based on the missing information principle, we have

$$\mathbf{I}_o(\boldsymbol{\theta}) = \mathbf{I}_c(\boldsymbol{\theta}) - \mathbf{I}_m(\boldsymbol{\theta}), \tag{29}$$

where $\mathbf{I}_c(\boldsymbol{\theta})$ and $\mathbf{I}_m(\boldsymbol{\theta})$ are defined as follows:

$$\mathbf{I}_c(\boldsymbol{\theta}) = E\left(-\frac{\partial^2 l_c}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}|y_{1:n}\right), \tag{30}$$

$$\mathbf{I}_m(\boldsymbol{\theta}) = E\left(\frac{\partial l_c}{\partial\boldsymbol{\theta}}\frac{\partial l_c}{\partial\boldsymbol{\theta}^T}|y_{1:n}\right) - E\left(\frac{\partial l_c}{\partial\boldsymbol{\theta}}|y_{1:n}\right)E\left(\frac{\partial l_c}{\partial\boldsymbol{\theta}^T}|y_{1:n}\right). \tag{31}$$

The first-order derivatives of $l_c(\boldsymbol{\theta})$ are given by

$$\frac{\partial l_c}{\partial\omega} = \frac{1}{\omega}\sum_{t=1}^{n} u_t - \frac{1}{1-\omega}\sum_{t=1}^{n}(1-u_t), \tag{32}$$

$$\frac{\partial l_c}{\partial\beta} = \sum_{t=1}^{n}(1-u_t)\left\{y_t - n_t\frac{\exp\left(\mathbf{x}_t^T\beta + z_t\right)}{1+\exp\left(\mathbf{x}_t^T\beta + z_t\right)}\right\}\mathbf{x}_t, \tag{33}$$

$$\frac{\partial l_c}{\partial\phi} = \frac{1}{\sigma^2}\sum_{t=1}^{n}\left(z_t - \phi^T\mathbf{s}_{t-1}\right)\mathbf{s}_{t-1}, \tag{34}$$

$$\frac{\partial l_c}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{t=1}^{n}\left(z_t - \phi^T \mathbf{s}_{t-1}\right)^2. \tag{35}$$

The second-order derivatives of $l_c(\theta)$ are given by

$$\frac{\partial^2 l_c}{\partial \omega \partial \omega} = -\frac{1}{\omega^2}\sum_{t=1}^{n} u_t - \frac{1}{(1-\omega)^2}\sum_{t=1}^{n}(1-u_t), \tag{36}$$

$$\frac{\partial^2 l_c}{\partial \beta \partial \beta^T} = -\sum_{t=1}^{n}(1-u_t)n_t \frac{\exp\left(\mathbf{x}_t^T\beta + z_t\right)}{\left[1+\exp\left(\mathbf{x}_t^T\beta + z_t\right)\right]^2}\mathbf{x}_t\mathbf{x}_t^T, \tag{37}$$

$$\frac{\partial^2 l_c}{\partial \phi \partial \phi^T} = -\frac{1}{\sigma^2}\sum_{t=1}^{n}\mathbf{s}_{t-1}\mathbf{s}_{t-1}^T, \tag{38}$$

$$\frac{\partial^2 l_c}{\partial \sigma \partial \sigma} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{t=1}^{n}\left(z_t - \phi^T \mathbf{s}_{t-1}\right)^2, \tag{39}$$

$$\frac{\partial^2 l_c}{\partial \phi \partial \sigma} = -\frac{2}{\sigma^3}\sum_{t=1}^{n}\left(z_t - \phi^T \mathbf{s}_{t-1}\right)\mathbf{s}_{t-1}. \tag{40}$$

Again, particle filtering and smoothing techniques are used to approximate the conditional expectations in $\mathbf{I}_c(\boldsymbol{\theta})$ and $\mathbf{I}_m(\boldsymbol{\theta})$.

In principle, the variance/covariance matrix can be approximated by taking the inverse of the observed information matrix. However, the computation of the inverse is often problematic. As indicated by Kim and Stoffer [25], the observed information matrix is not guaranteed to be numerically positive definite. To address this problem, we slightly modify Louis's formula by introducing a slack variable ξ, such that

$$\mathbf{I}_o(\boldsymbol{\theta}) = \mathbf{I}_c(\boldsymbol{\theta}) - (1-\xi)\mathbf{I}_m(\boldsymbol{\theta}), \tag{41}$$

where ξ is a non-negative variable ranging from 0 to 1. In practice, we can iteratively increase this value until the observed information matrix can be inverted.

3.3. Particle methods

Particle filtering [21] and particle smoothing [22] belong to the class of sequential Monte Carlo (SMC) methods [28]. These particle methods can be viewed as the non-linear and non-Gaussian extensions of the popular Kalman filtering and smoothing algorithms for traditional state-space models. Rather than yielding a single estimate for the filter or the smoother, as computed through conventional Kalman filtering and smoothing, particle methods provide a set of particles with associated weights to approximate the conditional densities governing the filters and smoothers. Implemented via sequential importance sampling (SIS), in the E-step of the EM algorithm, particle methods provide approximate solutions to the intractable integrals corresponding to the conditional expectations of functions of the latent components given the

observed data. However, sample degeneracy is a typical problem for SIS methods. In particular, degeneracy occurs when particles have small weights or even negative weights, rendering their contributions to the conditional density negligible. Resampling (e.g., bootstrapping) offers a recourse for eliminating particles with negligible effects. Kim [29] provides an elegant treatment of particle filtering and smoothing for state-space models.

Particle filtering

For the parameter-driven ZIB model, we implement particle filtering by first generating $\mathbf{s}_{0|0}^{(i)} \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. Then for $t = 1, \ldots, n$:

(F.1) Generate $\mathbf{s}_{t|t-1}^{(i)} \sim \mathcal{N}_p\left(\boldsymbol{\Phi}\mathbf{s}_{t-1|t-1}^{(i)}, \boldsymbol{\Sigma}\right)$ and $u_{t|t-1}^{(i)} \sim \text{Bernoulli}(\omega)$.

(F.2) Compute the filtering weights

$$q_{t|t-1}^{(i)} \propto \binom{n_t}{y_t}\left(\left(1-u_{t|t-1}^{(i)}\right)\pi_{t|t-1}^{(i)}\right)^{y_t}\left(1-\left(1-u_{t|t-1}^{(i)}\right)\pi_{t|t-1}^{(i)}\right)^{n_t-y_t}, \tag{42}$$

where $\text{logit}\left(\pi_{t|t\text{-}1}^{(i)}\right) = \mathrm{x}_t^{\mathrm{T}}\beta + z_{t|t\text{-}1}^{(i)}$ and $z_{t|t-1}^{(i)}$ is the first element of $\mathbf{s}_{t|t-1}^{(i)}$.

(F.3) Generate $\left(\mathbf{s}_{t|t}^{(i)}, u_{t|t}^{(i)}\right)$ by resampling $\left(\mathbf{s}_{t|t-1}^{(i)}, u_{t|t-1}^{(i)}\right)$ with replacement based on the preceding filtering weights.

As a byproduct of the particle filtering, the observed-data log-likelihood can be approximated by

$$\sum_{t=1}^{n} \log\left(\frac{1}{N}\sum_{i=1}^{N} q_{t|t-1}^{(i)}\right), \tag{43}$$

where N is the number of particles in the filtering step.

Particle smoothing

Next, we employ the particle smoothing algorithm proposed by Godsill et al. [22] to obtain the conditional expectations of the functions of the latent variables given the complete set of observed data. In this step, we first choose $\left(\mathbf{s}_{n|n}^{(r)}, u_{n|n}^{(r)}\right) = \left(\mathbf{s}_{n|n}^{(i)}, u_{n|n}^{(i)}\right)$ with probability $q_{n|n-1}^{(i)}$. Then for $t = n-1, \ldots, 1$:

(S.1) Calculate the smoothing weights

$$q_{t|n}^{(i)} \propto q_{t|t-1}^{(i)} \exp\left\{-\frac{1}{2\sigma^2}\left(z_{t+1|n}^{(i)} - \phi^T\mathbf{s}_{t|t}^{(i)}\right)^2\right\}\omega^{u_{t+1|n}^{(i)}}(1-\omega)^{1-u_{t+1|n}^{(i)}}. \tag{44}$$

(S.2) Choose $\left(\mathbf{s}_{t|n}^{(r)}, u_{t|n}^{(r)}\right) = \left(\mathbf{s}_{t|t}^{(i)}, u_{t|t}^{(i)}\right)$ with probability $q_{t|n}^{(i)}$.

We obtain independent realizations by repeating the preceding process for $r = 1, \ldots, R$.

4. Simulation studies

In this section, we investigate through simulation two salient issues pertaining to the proposed modeling frameworks. In the first part, we explore the convergence of the MCEM algorithm through simulated examples, and investigate the finite sample distributional properties of the parameter estimators through a comprehensive simulation study. In the second part, we present a simulation study to compare the performance of the proposed ZIB models to their counterpart ZIP models in characterizing zero-inflated count time series.

4.1. Evaluation of the MCEM algorithm

We consider time series data simulated from four different parameter-driven models: ZIB + AR (2), binomial + AR(2), ZIB + AR(1), and binomial + AR(1). The sample size is set to 300 and the number of cases n_t for each time point is set to 30. All of the models feature the following linear predictor:

$$\text{logit}(\pi_t) = \beta_0 + \beta_1 x_{1,t} + z_t, \tag{45}$$

where $x_{1,t}$ is a covariate series generated from a standard uniform distribution. The true parameters for the most complicated model ZIB + AR(2) are as follows:

$$\omega = 0.3, \beta_0 = 2, \beta_1 = -3, \phi_1 = 0.8, \phi_2 = -0.6, \text{and } \sigma = 0.5. \tag{46}$$

For the rest of the models considered, the corresponding parameters are set to 0 if no such a form is included. Autoregressive (AR) coefficients are chosen to assure stationarity of the series. In fitting the models, the number of particle filters (N) is set to 500 and the number of particle smoothers (R) is set to 300. We stop the MCEM algorithm after 300 iterations. **Table 1** presents the parameter estimates for the simulated data corresponding to the four parameter-driven models.

Figure 1 shows the trace plots of the log-likelihood for the four fitted parameter-driven models. Note that the log-likelihood of the MCEM algorithm is not strictly increasing at each iteration due to the introduction of Monte Carlo errors. However, the log-likelihood stabilizes after a few dozen iterations with slight fluctuations around the maximal value. **Figure 2** shows the trace plots for the parameter estimates from the most complex fitted model, ZIB + AR(2). The plots indicate that the parameter estimates converge to the MLEs quickly with negligible

	ω	β_0	β_1	ϕ_1	ϕ_2	σ
True	0.300	2.000	−3.000	0.800	−0.600	0.500
Binomial + AR(1)		1.984	−2.968	0.800		0.540
ZIB + AR(1)	0.283	2.124	−2.930	0.781		0.563
Binomial + AR(2)		1.989	−3.012	0.852	−0.620	0.499
ZIB + AR(2)	0.293	1.992	−2.872	0.831	−0.576	0.506

Table 1. True and estimated parameters for the simulated examples.

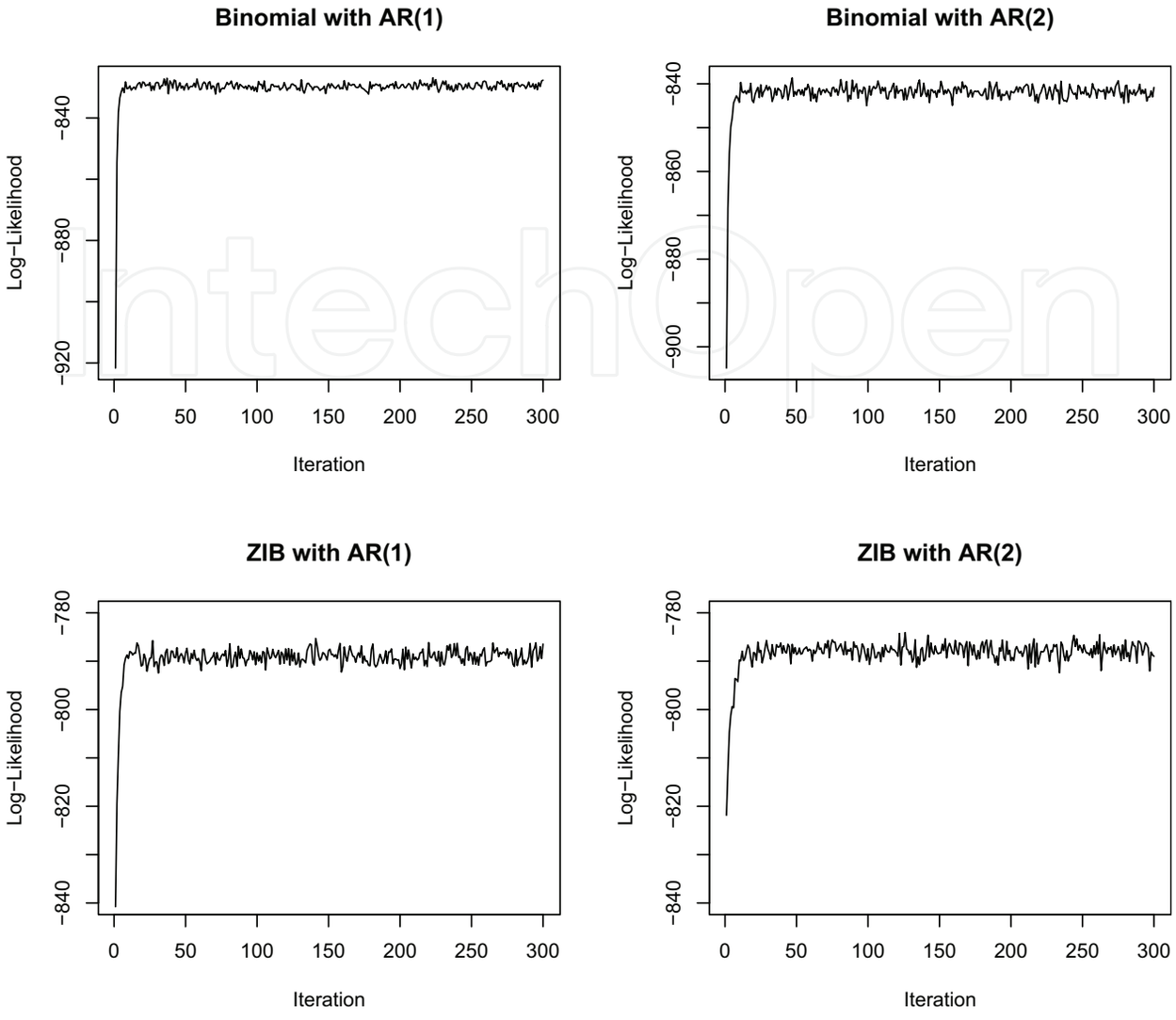


Figure 1. Trace plots of the log-likelihood for fitted parameter-driven models based on simulated data.

fluctuations. The trace plots of the parameter estimates for the other three models exhibit similar patterns (results not shown). In practice, we recommend always checking the trace plots of the estimates to assess convergence of the MCEM algorithm.

We next investigate the finite sample distributional properties of the parameter estimators from the MCEM algorithm. We consider the same parameter-driven models presented in the preceding simulated example. For each model structure, 500 replications are generated based on sample sizes of 200 and 500. We employ the proposed MCEM algorithm to fit models based on these replications, and record the subsequent parameter estimates and their standard errors. As the MECM algorithm is computationally expensive, we set the number of particles for both filters and smoothers to 200, and the stopping iteration for the MCEM algorithm at 100. In **Tables 2**–**3**, we provide the simulation results based on the most complex model, ZIB + AR(2).

In general, the mean and median of the estimates converge to the true parameters, with a minor degree of negative bias associated with the estimation of the AR coefficients. The empirical

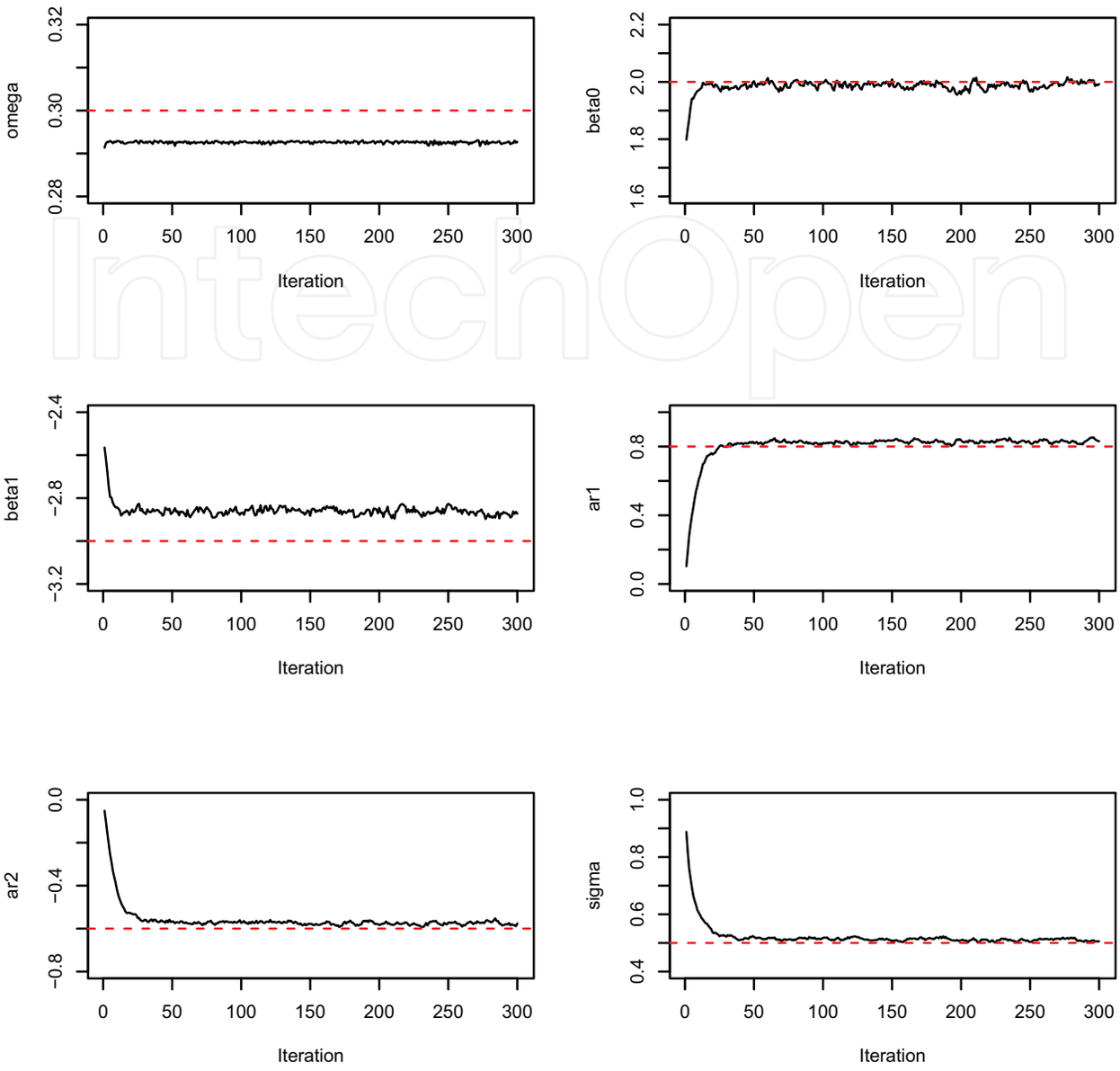


Figure 2. Trace plots of the estimated parameters for the fitted ZIB + AR(2) model.

standard deviations (ESDs) are reasonably close to the average asymptotic standard errors (ASEs). Therefore, the standard errors calculated by Louis's method prove to be sufficient. As the sample size increases from 200 to 500, the bias for the estimation of the AR coefficients attenuates, and the standard errors tend to diminish. The two behaviors indicate that weak convergence holds. The results for the other three parameter-driven models are analogous to those presented in **Tables 2–3**. **Tables 4–9** show the simulation results for the binomial + AR(2) model, ZIB + AR(1) model, and binomial + AR(1) model, respectively.

The normality of the parameter estimators is assessed by Q-Q plots based on the sets of replicated estimates (figures not shown). For the most complex ZIB + AR(2) model, approximate normality holds for the finite sample distribution of the parameter estimators, with

http://dx.doi.org/10.5772/intechopen.71336

	True	Mean	Median	ESD	ASE
ω	0.300	0.299	0.295	0.032	0.032
β_0	2.000	1.999	1.992	0.139	0.166
β_1	−3.000	−2.992	−2.980	0.235	0.224
ϕ_1	0.800	0.743	0.757	0.120	0.165
ϕ_2	−0.600	−0.563	−0.572	0.104	0.145
σ	0.500	0.504	0.508	0.063	0.098

Table 2. Summary statistics for replicated parameter estimates from fitted ZIB + AR(2) models with sample size 200.

	True	Mean	Median	ESD	ASE
ω	0.300	0.300	0.299	0.020	0.020
β_0	2.000	2.002	2.002	0.081	0.104
β_1	−3.000	−3.006	−3.007	0.138	0.139
ϕ_1	0.800	0.754	0.754	0.076	0.095
ϕ_2	−0.600	−0.566	−0.573	0.067	0.086
σ	0.500	0.509	0.510	0.039	0.057

Table 3. Summary statistics for replicated parameter estimates from fitted ZIB + AR(2) models with sample size 500.

	True	Mean	Median	ESD	ASE
β_0	2.000	1.998	2.003	0.108	0.167
β_1	−3.000	−3.002	−3.010	0.179	0.174
ϕ_1	0.800	0.783	0.783	0.087	0.101
ϕ_2	−0.600	−0.593	−0.596	0.080	0.094
σ	0.500	0.496	0.494	0.052	0.062

Table 4. Summary statistics for replicated parameter estimates from fitted binomial + AR(2) models with sample size 200.

	True	Mean	Median	ESD	ASE
β_0	2.000	1.995	1.989	0.070	0.101
β_1	−3.000	−2.994	−2.995	0.113	0.108
ϕ_1	0.800	0.791	0.791	0.057	0.063
ϕ_2	−0.600	−0.593	−0.595	0.053	0.059
σ	0.500	0.498	0.496	0.032	0.038

Table 5. Summary statistics for replicated parameter estimates from fitted binomial + AR(2) models with sample size 500.

	True	Mean	Median	ESD	ASE
ω	0.300	0.299	0.299	0.031	0.032
β_0	2.000	1.971	1.971	0.208	0.251
β_1	−3.000	−2.982	−2.969	0.199	0.210
ϕ_1	0.800	0.763	0.770	0.073	0.067
σ	0.500	0.500	0.502	0.056	0.063

Table 6. Summary statistics for replicated parameter estimates from fitted ZIB + AR(1) models with sample size 200.

	True	Mean	Median	ESD	ASE
ω	0.300	0.299	0.299	0.020	0.021
β_0	2.000	1.984	1.989	0.135	0.168
β_1	−3.000	−2.992	−2.991	0.133	0.132
ϕ_1	0.800	0.781	0.785	0.041	0.040
σ	0.500	0.500	0.499	0.035	0.040

Table 7. Summary statistics for replicated parameter estimates from fitted ZIB + AR(1) models with sample size 500.

	True	Mean	Median	ESD	ASE
β_0	2.000	2.006	2.024	0.192	0.233
β_1	−3.000	−2.987	−2.988	0.165	0.167
ϕ_1	0.800	0.782	0.788	0.054	0.056
σ	0.500	0.497	0.496	0.051	0.052

Table 8. Summary statistics for replicated parameter estimates from fitted binomial + AR(1) models with sample size 200.

	True	Mean	Median	ESD	ASE
β_0	2.000	1.997	1.996	0.125	0.168
β_1	−3.000	−2.997	−2.994	0.106	0.106
ϕ_1	0.800	0.787	0.789	0.035	0.035
σ	0.500	0.499	0.499	0.030	0.033

Table 9. Summary statistics for replicated parameter estimates from fitted binomial + AR(1) models with sample size 500.

slightly non-normal tail behavior (thick or thin) evident for the estimated AR coefficients. As the sample size is increased from 200 to 500, this non-normal behavior is attenuated. Similar patterns are observed for the other three parameter-driven models.

4.2. Model comparison

As previously mentioned, based on a Poisson mixture distribution, extensive methodology has been published to deal with count time series with excess zeros. In addition, the Poisson distribution provides an accurate approximation to the binomial distribution when the sample size is large and the success probability is small. Therefore, one may question whether Poisson-type models are sufficient for approximating binomial-type models when data are generated from a binomial mixture distribution. In this section, we try to address this question through a simulation study.

Two different types of ZIB models are proposed in this work: the parameter-driven ZIB model, and the observation-driven ZIB model. To evaluate the propriety of the binomial-type models, we consider two corresponding Poisson-type counterparts: the parameter-driven ZIP model, and the observation-driven ZIP model. We assess the performance of the four models under two scenarios: first, where data are generated from the parameter-driven ZIB model, and second, where data are generated from the observation-driven ZIB model.

To denote the parameter-driven ZIB/ZIP model with an AR(p) latent process, we use PDZIB(p)/PDZIP(p). Similarly, we use ODZIB(p)/ODZIP(p) to denote the observation-driven ZIB/ZIP model with p lagged responses employed as covariates.

In the first scenario, data are generated from a PDZIB(2) model having the same form as that provided in Section 4.1. To reduce the computational burden associated with fitting the models, 100 replicated series of length 200 are generated. We fit four different zero-inflated models to each of the series. For the two parameter-driven models, we specify a latent autoregressive process of order two, and employ the MECM algorithm to fit the models. For the two observation-driven models, we incorporate the lagged responses y_{t-1} and y_{t-2} to account for the temporal correlation, and employ the Newton–Raphson algorithm to fit the models.

In the second scenario, data are generated from an ODZIB(2) model featuring the following structures:

$$\text{logit}(\pi_t) = \beta_0 + \beta_1 x_{1,t} + \phi_1 y_{t-1} + \phi_2 y_{t-2}, \text{ and } \text{logit}(\omega) = \gamma_0. \tag{47}$$

Here, $x_{1,t}$ is a covariate series generated from a standard uniform distribution, and ϕ_1 and ϕ_2 are the autoregressive coefficients for the lagged responses y_{t-1} and y_{t-2}, respectively. The values of the true parameters are the same as those for the parameter-driven model.

Again, we generate 100 replications of length 200 based on the preceding model. The same four zero-inflated models are fit to each of the replications. The Akaike information criterion (AIC) [30] is used to guide the selection of an optimal model in both scenarios. To evaluate the magnitude of the absolute difference in AIC values, Burnham and Anderson [31] provide the following guidelines (**Table 10**).

Thus, a difference in AIC values of two or more is considered meaningful, and a difference of 10 or more is considered pronounced.

Difference in AIC	Level of empirical support for model with larger AIC
0–2	Substantial
4–7	Considerably less
>10	Essentially none

Table 10. Guidelines for assessing AIC differences.

Figure 3 illustrates the performance of the four zero-inflated models, in terms of AIC differences, when data are generated from a PDZIB(2) model. The PDZIB(2) model serves as the reference for model comparison. Each point represents the difference in the AIC value between the target model and the reference model. As evident from the figure, the PDZIB(2) model markedly outperforms the other three models for all 100 replications, with AIC differences over 50. Although vastly inferior to the PDZIB(2) model, the PDZIP(2) model performs better than the two observation-driven models. The ODZIB(2) performs the worst among the four models considered. Parameter-driven models clearly exhibit a substantial advantage over observation-driven models when the underlying data arise via a parameter-driven approach.

Figure 4 shows the performance of the four zero-inflated models, in terms of AIC differences, when data are generated from an ODZIB(2) model. Similarly, the ODZIB(2) model serves as

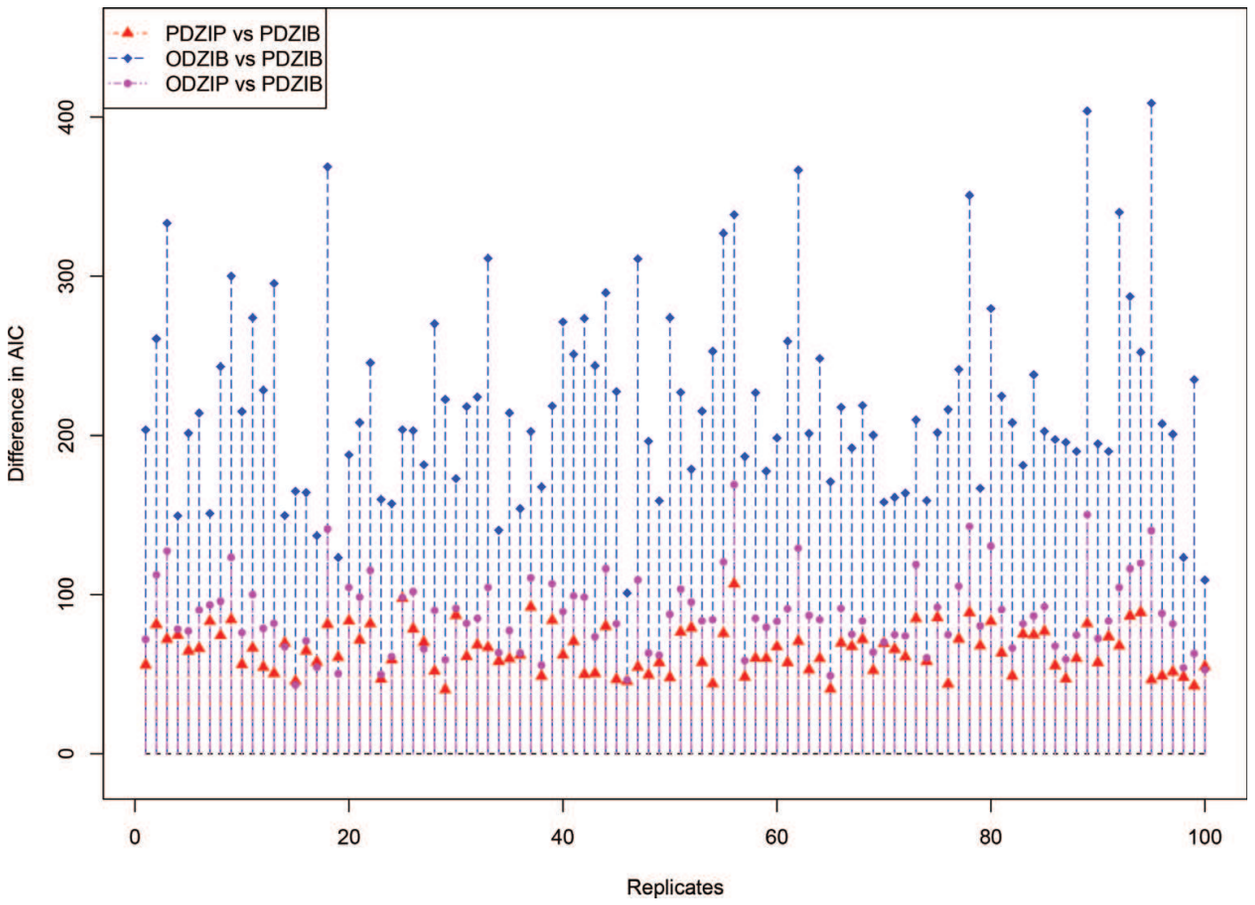


Figure 3. AIC differences of zero-inflated fitted models relative to parameter-driven ZIB fitted models.

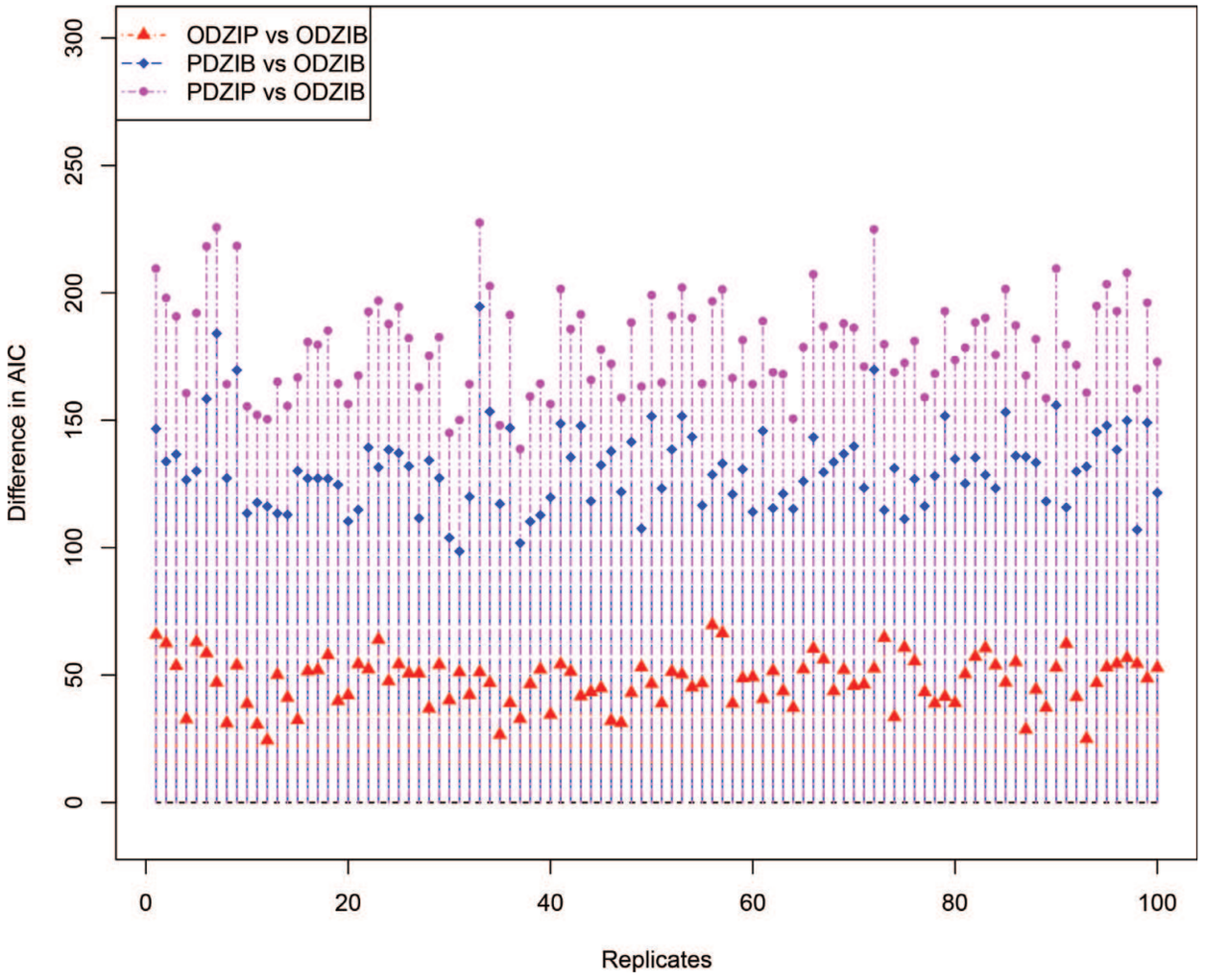


Figure 4. AIC differences of zero-inflated fitted models relative to observation-driven ZIB fitted models.

the reference. The ODZIB(2) model easily performs the best among all four models for all 100 replications, reflecting a substantial improvement in model fit over the other three models based on AIC differences (>20). Compared to the two parameter-driven models, the ODZIP(2) model accommodates the data much more appropriately. Between the two parameter-driven models, the PDZIB(2) model is substantially favored over the PDZIP(2) model. Thus, observation-driven models markedly outperform parameter-driven models when the underlying data arise via an observation-driven approach.

We close this section with a brief discussion of issues germane to model selection. These issues are relevant not only in evaluating the results of the preceding simulations, but also in facilitating the choice of a model in practice.

First, one may question which class of models should be considered when coping with binomial time series data with excess zeros. In the simulation sets, the fitted parameter-driven models markedly outperform the fitted observation-driven models when data are generated via a parameter-driven approach. Although parameter-driven models are computationally expensive to fit, observation-driven models do not appear to provide an adequate characterization of the

data in such settings. Additionally, unlike observation-driven models, parameter-driven models provide a description of the underlying latent processes that govern the temporal correlation and zero inflation. Observation-driven models, in contrast, outperform parameter-driven models when the underlying data are generated via an observation-driven approach. In general, the selection of the class of models depends on the conceptualization of the model structure and the perceived value of recovering and investigating the underlying latent processes. However, in the context of zero-inflated count time series, since an understanding of the phenomenon that gives rise to the data will rarely inform the practitioner as to whether the parameter-driven or observation-driven conceptualization is more appropriate, we recommend the use of AIC or an alternate likelihood-based selection criterion in choosing between these two model classes.

Second, one may question which distribution should be used when dealing with count time series with excess zeros. The Poisson-type model with an offset is often considered an appropriate approximating model for a binomial-type model when the sample size is large and the success probability is low. However, in the presence of zero inflation, our simulation results indicate the necessity of using binomial-type models over their Poisson counterparts when the underlying distribution is actually a binomial mixture. In practice, if the dynamics of the phenomenon that gives rise to the data do not inform the underlying data generating distribution, we again recommend the use of AIC or another likelihood-based criterion in choosing an appropriate distribution.

5. Application

In this section, to illustrate our proposed methodology, we consider an application pertaining to the diagnosis coding of a severe disease, Kaposi's sarcoma (KS). The application concerns the assessment of a particular level change for a primary KS diagnosis. The data used are extracted from the Healthcare Cost and Utilization Project (HCUP) database. We identify all hospitalizations during the period from January 1998 through December 2011 during which a primary or secondary diagnosis of KS is received. For case ascertainment, we use the International Classification of Diseases, 9^{th} Revision, Clinical Modification (ICD-9-CM), code 176. We then aggregate all cases of KS by month to produce a national sample of the monthly KS hospitalizations. The data consist of monthly counts of both primary and overall KS hospitalizations from January 1998 to December 2011. The sample size for both KS series is 168. **Figure 5** shows both the primary KS count time series and the overall KS count time series. In the latter, the overall KS count serves as the denominator for the binomial-type model and the offset for the Poisson-type model.

A coding change was implemented in early 2008, during which many hospitals may have modified the coding convention by switching the primary code to secondary, as this modification may lead to an increase in hospital reimbursements. During the study period, a large number of zero counts is observed and data among adjacent points seem to be highly correlated. Since the primary KS count series exhibits a relatively large degree of zero-inflation (appropriately 25% of the values are zero), we apply our proposed ZIB models to characterize the data.

http://dx.doi.org/10.5772/intechopen.71336

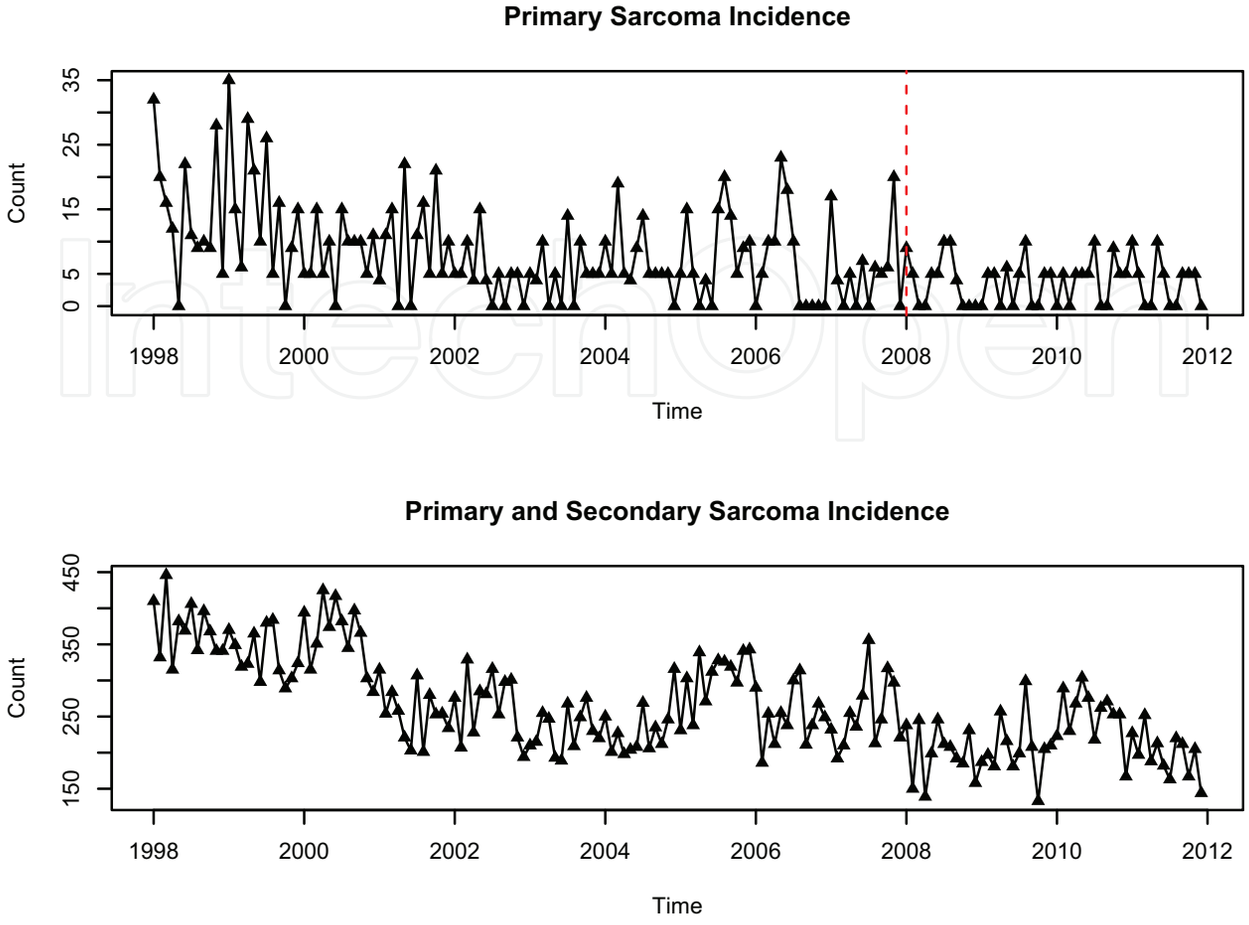


Figure 5. Monthly time series plots of primary KS hospitalizations (top panel) and overall KS hospitalizations (bottom panel) from January 1998 to December 2011.

Our analysis focuses on two objectives. First, we aim to model the dynamic pattern of the primary KS series; in particular, we are interested in determining the appropriate order of the autoregressive process embedded in the series, and evaluate whether there is a significant level change at January 2008. Second, we aim to compare the performance of our proposed ODZIB(p) and PDZIB(p) models to their counterpart ODZIP(p) and PDZIP(p) models.

For potential autocorrelation structures, we let p be either 1 or 2. As a result, we consider eight candidate models in total. Each of the models features an indicator to represent an intervention in January 2008, which allows us to test whether there is significant level change at this time period.

Specifically, for the two PDZIB(p) models, we employ the following linear predictor:

$$\text{logit}(\pi_t) = \beta_0 + \beta_1 x_t + z_t, \tag{48}$$

$$z_t = \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t, \tag{49}$$

where t is a discrete time index, and $x_t = I_{(t>2008)}$ is a dummy variable indicating whether the index t is greater than the predefined change point (January 2008). Thus, β_1 reflects the level

Model	AIC	ω	β_0	β_1	ϕ_1	ϕ_2	σ
PDZIB(1)	922.98	0.248	−3.349	−0.249	−0.223		0.430
		(0.034)	(0.051)	(0.120)	(0.160)		(0.044)
PDZIP(1)	923.31	0.248	−3.389	−0.242	−0.241		0.410
		(0.034)	(0.051)	(0.116)	(0.166)		(0.043)
ODZIB(1)	1039.80	0.341	−3.184	−0.319	−0.007		
		(0.061)	(0.024)	(0.086)	(0.002)		
ODZIP(1)	1030.04	0.341	−3.224	−0.309	−0.007		
		(0.061)	(0.046)	(0.084)	(0.004)		
PDZIB(2)	922.98	0.248	−3.359	−0.237	−0.120	0.264	0.426
		(0.034)	(0.054)	(0.126)	(0.166)	(0.153)	(0.046)
PDZIP(2)	924.09	0.248	−3.395	−0.230	−0.119	0.263	0.402
		(0.034)	(0.052)	(0.118)	(0.178)	(0.158)	(0.045)
ODZIB(2)	1038.11	0.341	−3.250	−0.275	−0.008	0.007	
		(0.061)	(0.033)	(0.088)	(0.002)	(0.002)	
ODZIP(2)	1028.49	0.341	−3.288	−0.266	−0.007	0.007	
		(0.061)	(0.058)	(0.087)	(0.004)	(0.004)	

Table 11. Model fitting results for eight different zero-inflated models.

change in KS counts due to the coding practice, and the ϕ_i denote the coefficients for the autoregressive process.

For the two ODZIB(p) models, we employ the following linear predictor:

$$\text{logit}\,(\pi_t) = \beta_0 + \beta_1 x_t + \sum_{i=1}^{p} \phi_i y_{t-i}, \tag{50}$$

where β_1 and ϕ_i reflect parameters analogous to those defined for the parameter-driven setting.

In addition, we consider four comparable Poisson-type models based on the work by Yang et al. [5, 7]. For the two PDZIP(p) models, we employ the linear predictor

$$\log\left(\mu_t\right) = \log\left(n_t\right) + \beta_0 + \beta_1 x_t + z_t, \tag{51}$$

$$z_t = \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t. \tag{52}$$

For the two ODZIP(p) models, we employ the linear predictor

$$\log\left(\mu_t\right) = \log\left(n_t\right) + \beta_0 + \beta_1 x_t + \sum_{i=1}^{p} \phi_i y_{t-i}. \tag{53}$$

Here, n_t serves as an offset variable representing the overall number of KS diagnoses. AIC is used to guide the selection of the optimal model.

Table 11 features results for the eight fitted candidate models. The parameter estimates along with their standard errors are presented. All eight models indicate a significant level change for the primary KS series after the introduction of the potential coding change practice ($\beta_1 < 0$). Among the first four models, which feature an autocorrelation structure of order one, parameter-driven models are deemed superior to observation-driven models, with AIC differences over 100. The PDZIB(1) model is slightly favored over the PDZIP(1) in terms of the AIC value. We observe similar patterns in the last four models, which feature an autocorrelation structure of order two. Among the parameter-driven models, adding a second order to the autocorrelation offers little improvement in model fit, since the increase in goodness-of-fit is offset by a decrease in parsimony. Therefore, the best model appears to be PDZIB(1).

6. Conclusion

Count time series featuring a preponderance of zeros are commonly encountered in a variety of scientific applications. In characterizing such series, modeling frameworks that assume a Poisson mixture distribution have been extensively studied. However, minimal work has been focused on modeling frameworks that assume a binomial mixture distribution. When data are more naturally assumed to arise from the latter, a Poisson-type model with an offset is often employed; however, the propriety of such an approximation is unclear.

We propose two general classes of models to effectively characterize a count time series that arises from a zero-inflated binomial mixture distribution. The observation-driven ZIB model, formulated in the partial likelihood framework, is fit using the Newton–Raphson algorithm. The parameter-driven ZIB model, formulated in the state-space framework, is fit using the MCEM algorithm. When data are generated from a binomial mixture, our proposed ZIB models outperform their Poisson-type counterparts. We illustrate our methodology with an application that assesses a particular level change for a diagnosis code.

Future work involves extending the current frameworks to the zero-inflated beta-binomial (ZIBB) model. Both observation-driven and parameter-driven ZIBB models can be formulated and fit based on methodological developments similar to those presented in this work. However, weak identifiability could arise as a potentially problematic issue in fitting the parameter-driven ZIBB model, as not only the overdispersion explicitly induced by the beta distribution but also the correlated random effects account for any excess variability in the data [5]. In addition, we could consider more complicated correlation structures by incorporating moving average components in the linear predictors for parameter-driven models. Such an extension necessitates non-trivial revisions to the state-space model formulation and the complete-data likelihood, which warrant further investigation.

Author details

Fan Tang[1] and Joseph E. Cavanaugh[2]*

*Address all correspondence to: joe-cavanaugh@uiowa.edu

1 Genentech Inc., South San Francisco, CA, USA

2 Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA, USA

References

[1] Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992;**34**:1-14

[2] Yau KKW, Lee AH. Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. Statistics in Medicine. 2001;**20**:2907-2920

[3] Min Y, Agresti A. Random effect models for repeated measures of zero-inflated count data. Statistical Modeling. 2005;**5**:1-19

[4] Yau KKW, Lee AH, Carrivick PJW. Modeling zero-inflated count series with application to occupational health. Computer Methods and Programs in Biomedicine. 2004;**74**:47-52

[5] Yang M, Zamba GKD, Cavanaugh JE. State-space models for count time series with excess zeros. Statistical Modeling. 2015;**15**:70-90

[6] Hall DB. Zero-inflated Poisson and binomial regression with random effects: A case study. Biometrics. 2000;**56**:1030-1039

[7] Yang M, Zamba GKD, Cavanaugh JE. Markov regression models for count time series with excess zeros: A partial likelihood approach. Statistical Methodology. 2013;**14**:26-38

[8] Cox DR. Statistical analysis of time series: Some recent developments. Scandinavian Journal of Statistics. 1981;**8**:93-115

[9] Zeger SL, Qaqish B. Markov regression models for time series: A quasi-likelihood approach. Biometrics. 1988;**44**:1019-1031

[10] Slud E, Kedem B. Partial likelihood analysis of logistic regression and autoregression. Statistica Sinica. 1994;**4**:89-106

[11] Davis RA, Dunsmuir WTM, Streett SB. Observation-driven models for Poisson counts. Biometrika. 2003;**90**:777-790

[12] Freeland RK, McCabe BPM. Analysis of low count time series data by Poisson autoregression. Journal of Time Series Analysis. 2004;**25**:701-722

[13] Fokianos K, Kedem B. Partial likelihood inference for time series following generalized linear models. Journal of Time Series Analysis. 2004;**25**:173-197

[14] Hudecoa Š. Structural changes in autoregressive models for binary time series. Journal of Statistical Planning and Inference. 2013;**143**:1744-1752

[15] Zeger SL. A regression model for time series of counts. Biometrika. 1988;**75**:621-629

[16] Chan KS, Ledolter J. Monte Carlo EM estimation for time series models involving counts. Journal of the American Statistical Association. 1995;**90**:242-252

[17] Nelson KP, Leroux BG. Statistical models for autocorrelated count data. Statistics in Medicine. 2006;**25**:1413-1430

[18] Klingenberg B. Regression models for binary time series with gaps. Computational Statistics and Data Analysis. 2008;**52**:4076-4090

[19] Wu R, Cui Y. A parameter-driven logit regression model for binary time series. Journal of Time Series Analysis. 2014;**35**:462-477

[20] Kedem B, Fokianos K. Regression Models for Time Series Analysis. New Jersey: Wiley; 2002. p. 49-59

[21] Gordon NJ, Salmond DJ, Smith AFM. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings-F, Radar and Signal Processing. 1993;**140**:107-113

[22] Godsill SJ, Doucet A, West M. Monte Carlo smoothing for nonlinear time series. Journal of the American Statistical Association. 2004;**99**:156-168

[23] Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B. 1977;**39**:1-39

[24] Liu SI. Bayesian model determination for binary-time-series data with applications. Computational Statistics and Data Analysis. 2001;**36**:461-473

[25] Kim J, Stoffer DS. Fitting stochastic volatility models in the presence of irregular sampling via particle methods and the EM algorithm. Journal of Time Series Analysis. 2008;**29**:811-833

[26] Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society, Series B. 2009;**71**:319-392

[27] Louis TA. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society Series B. 1982;**44**:226-233

[28] Doucet A, Freitas ND, Gordon N. Sequential Monte Carlo Methods in Practice. New York: Springer; 2001

[29] Kim J. Parameter estimation in stochastic volatility models with missing data using particle methods and the EM algorithm. PhD Thesis. Pittsburgh, PA: University of Pittsburgh; 2005. p. 10-27

[30] Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974;**19**:716-723

[31] Burnham KP, Anderson DR. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd ed. New York: Springer; 2002. p. 70

IntechOpen

IntechOpen