Statistical models for count time series with excess zeros

Yang, Ming

Yang. (2012). Statistical models for count time series with excess zeros [University of Iowa].
https://doi.org/10.17077/etd.bcrq9mz0

STATISTICAL MODELS FOR COUNT TIME SERIES WITH EXCESS ZEROS

by

Ming Yang

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Biostatistics in the
Graduate College of The
University of Iowa

May 2012

Thesis Supervisors: Professor Joseph Cavanaugh
Assistant Professor Gideon Zamba

ABSTRACT

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.

In this thesis, we develop two classes of statistical models for zero-inflated time series. The first part of the thesis introduces a class of observation-driven models in a partial likelihood framework. Iterative algorithms (Newton-Raphson, Fisher Scoring, EM) are developed to obtain the maximum partial likelihood estimator (MPLE). We establish the asymptotic properties of the MPLE under certain regularity conditions. The performances of different partial-likelihood based model selection criteria are compared under model misspecification. In the second part of the thesis, we introduce a class of parameter-driven models in a state-space framework. To estimate the model parameters, we devise a Monte Carlo EM algorithm, where particle filtering and particle smoothing methods are employed to approximate the high-dimensional integrals in the E-step of the algorithm. Upon convergence, Louis' formula is used to find the observed information matrix.

The proposed models are illustrated with simulated data and an application based on public health surveillance for syphilis, a sexually transmitted disease (STD) that remains a major public health challenge in the United States. An R package, called ZIM (Zero-Inflated Models), has been developed to fit both observation-driven models and parameter-driven models.

Abstract Approved: ______________________

Thesis Supervisor

____

Title and Department

____

Date

____

Thesis Supervisor

____

Title and Department

____

Date

STATISTICAL MODELS FOR COUNT TIME SERIES WITH EXCESS ZEROS

by

Ming Yang

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Biostatistics in the
Graduate College of The
University of Iowa

May 2012

Thesis Supervisors: Professor Joseph Cavanaugh
Assistant Professor Gideon Zamba

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Ming Yang

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Biostatistics at the May 2012 graduation.

Thesis Committee: ______________________________
Joseph Cavanaugh, Thesis Supervisor

____
Gideon Zamba, Thesis Supervisor

____
Kung-Sik Chan

____
Michael Jones

____
Philip Polgreen

ACKNOWLEDGMENTS

First of all, I would like take this opportunity to thank all the faculty and my friends in the Department of Biostatistics. It has been one of the best decisions in my life to study at the University of Iowa.

Especially, I would like to thank my advisors, Drs. Joseph Cavanaugh and Gideon Zamba, for their excellent guidance and constant inspiration throughout the course of my Ph.D. study. I have received considerable help from them both in and outside the classroom. This work would be impossible without their tremendous support and encouragement.

I am also very grateful to all the other members of my thesis committee, Drs. Kung-Sik Chan, Michael Jones, and Philip Polgreen, for sharing their invaluable comments and suggestions.

Most importantly, I am deeply indebted to my parents and my brother for their support and love in my life. I thank them for understanding my life choices and encouraging me to be myself.

Finally, I would also like to thank all the coaches and student athletes at the University of Iowa. I bought season tickets for both football and basketball games in the last three years. It has been an awesome experience for me to be a loyal Hawkeye fan. Go Hawks!

ABSTRACT

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.

In this thesis, we develop two classes of statistical models for zero-inflated time series. The first part of the thesis introduces a class of observation-driven models in a partial likelihood framework. Iterative algorithms (Newton-Raphson, Fisher Scoring, EM) are developed to obtain the maximum partial likelihood estimator (MPLE). We establish the asymptotic properties of the MPLE under certain regularity conditions. The performances of different partial-likelihood based model selection criteria are compared under model misspecification. In the second part of the thesis, we introduce a class of parameter-driven models in a state-space framework. To estimate the model parameters, we devise a Monte Carlo EM algorithm, where particle filtering and particle smoothing methods are employed to approximate the high-dimensional integrals in the E-step of the algorithm. Upon convergence, Louis' formula is used to find the observed information matrix.

The proposed models are illustrated with simulated data and an application based on public health surveillance for syphilis, a sexually transmitted disease (STD) that remains a major public health challenge in the United States. An R package, called ZIM (Zero-Inflated Models), has been developed to fit both observation-driven models and parameter-driven models.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

1.1 Scientific Background

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 infections 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. Such zero-inflated time series also arise in hospital epidemiology where the focus is on the number of hospital-acquired infections in different intensive care units. Counts of rare infections typically contain an excess of zeros, and cannot be well accommodated by the widely used Poisson or negative binomial model. In addition, temporal correlation between adjacent observations is often present in data collected over time. Failure to account for zero-inflation and/or autocorrelation in the data may result in misleading inference and the detection of the spurious associations.

1.2 Literature Review

Regression models based on the zero-inflated Poisson (ZIP) distribution have been well developed for zero-inflated count data that are independently distributed (Lambert, 1992; Lee et al. 2001). To analyze repeated measures count data with a large number of zeros, Hall (2000) incorporated random effects in the classical ZIP regression model to account for within-subject correlation and between-subject heterogeneity. Marginal models (Hall & Zhang, 2004) and multi-level models (Lee et al., 2006) 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 by Van den Broek (1995), and was later extended to the setting where count data are correlated (Xiang et al., 2006). Despite the popularity of the ZIP models, the literature for count time series with excess zeros is sparse.

In general, there are two types of time series models: observation-driven models and parameter-driven models (Cox, 1981). These two types of models differ in the way they account for autocorrelation. In observation-driven models, the temporal correlation between adjacent observations is directly modeled through a function of past responses. In contrast, an unobserved latent process is employed in parameter-driven models to account for the serial correlation. Conditioning on the latent process, the observations are assumed to be independently distributed. Compared to observation-driven models, the concept behind parameter-driven models is more appealing. However, parameter estimation in parameter-driven models is often computationally burdensome (Chan & Ledolter, 1995; Durbin & Koopman, 2000). A comprehensive comparison of different estimation methods for Poisson parameter-driven models is presented by Nelson & Leroux (2006).

There is a large literature for count time series without zero-inflation, including both observation-driven models (Davis et al., 2003; Li, 1994; Zeger & Qaqish, 1988) and parameter-driven models (Chan & Ledolter, 1995; Oh & Lim 2001; Zeger 1988). Unfortunately, very few papers have been published for count time series with excess zeros. Yau et al. (2004) first presented a ZIP mixed autoregressive model and applied the model to evaluate a participatory ergonomics intervention in occupational health. The model they proposed belongs to the class of parameter-driven models. However, the first-order autoregressive structure they consider is too restrictive to accurately approximate the actual temporal correlation in many time series data.

1.3 Overview of the Thesis

In this thesis, we develop both observation-driven models and parameter-driven models for count time series with excess zeros. The zero-inflated models proposed here can be viewed as natural extensions of the existing time series models following generalized linear models. In the case when there is no zero-inflation in the data, our models will reduce to the traditional time series models based on the Poisson and negative binomial distributions.

The thesis is organized as follows. In the partial likelihood framework, we present a class of observation-driven models for zero-inflated time series in Chapter 2. Chapter 3 introduces a class of parameter-driven models in the state-space framework. In Chapter 4, we provide a general introduction to an R package that has been developed to implement the proposed methodologies. Chapter 5 concludes the thesis with a discussion of future directions.

CHAPTER 2
OBSERVATION-DRIVEN MODELS

In this chapter, our focus is on observation-driven models for zero-inflated time series. We first review partial likelihood inference for time series following generalized linear models (GLM). In the partial likelihood framework, we introduce a class of zero-inflated models for count time series with a large proportion of zeros. Iterative algorithms are derived to obtain the maximum partial likelihood estimator (MPLE). Under certain regularity conditions, we establish the asymptotic properties of the MPLE. We also compare the performances of different partial-likelihood based model selection criteria under model misspecification. Both simulated and real examples are presented to illustrate the proposed methodologies.

2.1 Partial Likelihood Inference

The concept of partial likelihood was originally developed by Cox (1972 & 1975) for time-to-event data through the Cox's proportional hazards model. However, the technique has also been used in a variety of other statistical frameworks since its introduction. For example, Kedem & Fokianos (2002, Chapters 1-4) combined partial likelihood theory with GLM methods to model time series data where the conditional distribution of the response series belongs to the exponential family. Such a modeling framework is very flexible and provides a unified approach for describing non-Gaussian time series which cannot be analyzed by the well-known autoregressive integrated moving average (ARIMA) models (Box & Jenkins, 1970). The basic idea behind partial likelihood is to simplify complicated likelihood functions by dropping certain unimportant components that have a minimal effect on the large-sample properties of the estimated parameters (e.g., consistency and asymptotic normality).

To illustrate the use of partial likelihood for temporally correlated data, we consider a pair of jointly distributed time series (X_t, Y_t) with $t = 1, \ldots, N$, where $\{Y_t\}$ is the response series and $\{X_t\}$ is the explanatory series. To simplify the notation, we let $X_{1:N} = (X_1, \ldots, X_N)$ and $Y_{1:N} = (Y_1, \ldots, Y_N)$. Based on the rules of conditional probability, the joint density of $X_{1:N}$ and $Y_{1:N}$ can be written as

$$f_{\boldsymbol{\theta}}(x_{1:N}, y_{1:N}) = f_{\boldsymbol{\theta}}(x_1) \prod_{t=2}^{N} f_{\boldsymbol{\theta}}(x_t | x_{1:t-1}, y_{1:t-1}) \prod_{t=1}^{N} f_{\boldsymbol{\theta}}(y_t | x_{1:t}, y_{1:t-1}), \tag{2.1}$$

where $\boldsymbol{\theta}$ is a set of unknown parameters. In the special case when $t = 1$, we define $f_{\boldsymbol{\theta}}(y_t | x_{1:t}, y_{1:t-1}) = f_{\boldsymbol{\theta}}(y_1 | x_1)$. According to Cox (1975), the second product on the right hand side of (2.1) constitutes a partial likelihood and can be employed for statistical inference about $\boldsymbol{\theta}$. The vector $\hat{\boldsymbol{\theta}}$ that maximizes the partial likelihood is called the MPLE. Under mild regularity conditions, the loss of information due to this simplification is negligible, and most of the large-sample properties still hold for the MPLE. On the other hand, the computational effort required for parameter estimation can be greatly reduced.

As mentioned earlier, the partial likelihood approach has already been employed by Kedem & Fokianos (2002) to model time series following GLMs. Specifically, they assume the conditional distribution of the response series $\{Y_t\}$ belongs to the following exponential family:

$$f_{Y_t}(y_t | \mathcal{F}_{t-1}) = \exp\left\{ \frac{y_t \theta_t - b(\theta_t)}{a_t(\phi)} + c(y_t; \phi) \right\}, \tag{2.2}$$

where θ_t and ϕ are the natural and dispersion parameters, respectively. Here $\mathcal{F}_{t-1}$ is a filtration that represents all information that is known at time $t-1$. Such a filtration typically includes past values of the response series, and past and possibly present values of the explanatory series. Similar to the GLM results for independent data, the conditional mean and variance of (2.2) are given by

$$\mu_t = \mathrm{E}(Y_t | \mathcal{F}_{t-1}) = b'(\theta_t)$$

and

$$\sigma_t^2 = \mathrm{Var}(Y_t|\mathcal{F}_{t-1}) = a_t(\phi)b''(\theta_t).$$

Typically we model the conditional mean of the response series $\{Y_t\}$ through the systematic component

$$g(\mu_t) = \eta_t = \mathbf{x}_{t-1}^\top \boldsymbol{\beta},$$

where $\mathbf{x}_{t-1} = (x_{t-1,1}, \ldots, x_{t-1,p})$ is a set of deterministic or random covariates and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is the vector of unknown parameters. Here, $g(\cdot)$ is a monotone link function used to relate the conditional mean μ_t to the linear predictor η_t. Although the exponential family is very general and includes many important distributions as special cases (e.g., Gaussian, binomial, and Poisson), it is not suitable for count data with a large number of zeros.

2.2 ZIP Autoregression

Based on the zero-inflated Poisson (ZIP) distribution, we introduce an autoregressive model for count time series with excess zeros. The proposed model is an extension of the Poisson autoregression discussed by Kedem & Fokianos (2002, Chapter 4). Let $\{Y_t\}$ denote the response series, composed of discrete count data. We assume the count series is conditionally distributed as ZIP(λ_t, ω_t), with probability mass function (p.m.f.) defined as follows:

$$f_{Y_t}(y_t|\mathcal{F}_{t-1}) = \omega_t I_{(y_t=0)} + (1-\omega_t)\exp(-\lambda_t)\lambda_t^{y_t}/y_t!, \tag{2.3}$$

or equivalently

$$f_{Y_t}(y_t|\mathcal{F}_{t-1}) = \begin{cases} \omega_t + (1-\omega_t)\exp(-\lambda_t), & \text{if } y_t = 0, \\ \\ (1-\omega_t)\exp(-\lambda_t)\lambda_t^{y_t}/y_t!, & \text{if } y_t > 0. \end{cases} \tag{2.4}$$

Here λ_t is the intensity parameter of the baseline Poisson distribution, and ω_t is often referred to as the zero-inflation parameter. To simplify the notation, we will use $y_{0,t}$ to represent $I_{(y_t=0)}$ for the remainder of the thesis.

In general, the ZIP distribution defined by (2.3) or (2.4) can be viewed as a two-component mixture of the Poisson distribution with a degenerate distribution having point mass at zero. Specifically, we assume there is a dichotomous variable u_t 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. For $s < t$, the variables u_t and u_s are assumed to be conditionally independent of each other, given $\mathcal{F}_{t-1}$.

We consider a hierarchical model where

$$u_t|\mathcal{F}_{t-1} \sim \text{Bernoulli}(\omega_t) \tag{2.5}$$

and

$$Y_t|u_t, \mathcal{F}_{t-1} \sim \text{Poisson}((1-u_t)\lambda_t). \tag{2.6}$$

It can be easily verified that the distribution of $Y_t|\mathcal{F}_{t-1}$ implied by (2.5) and (2.6) is identical to the ZIP distribution defined by (2.3) or (2.4). In the case when the zero-inflation parameter ω_t is equal to zero, the ZIP distribution simply reduces to the ordinary Poisson distribution.

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

$$\begin{aligned} F_{Y_t}(m|\mathcal{F}_{t-1}) &= \Pr(Y_t \leq m|\mathcal{F}_{t-1}) \\ &= \sum_{y_t=0}^{m} f_{Y_t}(y_t|\mathcal{F}_{t-1}) \\ &= \omega_t + (1-\omega_t)\exp(-\lambda_t)\sum_{y_t=0}^{m} \lambda_t^{y_t}/y_t!. \end{aligned}$$

Based on equations (2.5) and (2.6), the mean of $Y_t|\mathcal{F}_{t-1}$ can be expressed as

$$\begin{aligned}
\mathrm{E}(Y_t|\mathcal{F}_{t-1}) &= \mathrm{E}\{\mathrm{E}(Y_t|u_t, \mathcal{F}_{t-1})\} \\
&= \mathrm{E}\{(1-u_t)\lambda_t|\mathcal{F}_{t-1}\} \\
&= \lambda_t(1-\omega_t),
\end{aligned}$$

and the variance of $Y_t|\mathcal{F}_{t-1}$ can be written as

$$\begin{aligned}
\mathrm{Var}(Y_t|\mathcal{F}_{t-1}) &= \mathrm{E}\{\mathrm{Var}(Y_t|u_t, \mathcal{F}_{t-1})\} + \mathrm{Var}\{\mathrm{E}(Y_t|u_t, \mathcal{F}_{t-1})\} \\
&= \mathrm{E}\{(1-u_t)\lambda_t|\mathcal{F}_{t-1}\} + \mathrm{Var}\{(1-u_t)\lambda_t|\mathcal{F}_{t-1}\} \\
&= \lambda_t(1-\omega_t) + \lambda_t^2\omega_t(1-\omega_t) \\
&= \lambda_t(1-\omega_t)(1+\lambda_t\omega_t).
\end{aligned}$$

For the ZIP distribution, the variance-to-mean ratio or dispersion index

$$\mathrm{Var}(Y_t|\mathcal{F}_{t-1})/\mathrm{E}(Y_t|\mathcal{F}_{t-1}) = 1 + \lambda_t\omega_t$$

is greater than or equal to one. Thus, excess zeros in the data also result in overdispersion, which again cannot be accommodated by the ordinary Poisson distribution.

We now propose a ZIP autoregression in which the intensity parameter λ_t and zero-inflation parameter ω_t are modeled as follows:

$$\eta_t = \log \lambda_t = \mathbf{x}_{t-1}^\top \boldsymbol{\beta} \tag{2.7}$$

and

$$\xi_t = \mathrm{logit}(\omega_t) = \mathbf{z}_{t-1}^\top \boldsymbol{\gamma}, \tag{2.8}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ are the regression coefficients for the log-linear part (2.7) and logistic part (2.8), respectively. For convenience, we let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ denote the $(p+q)$-dimensional vector of unknown parameters. Here $\mathbf{x}_{t-1} = (x_{t-1,1}, \ldots, x_{t-1,p})^\top$ and $\mathbf{z}_{t-1} = (z_{t-1,1}, \ldots, z_{t-1,q})^\top$ denote vectors of past explanatory variables, into which functions of the lagged response series can be incorporated to account for serial correlation.

According to Kedem & Fokianos (2002), the general definition of partial likelihood is given by

$$\mathrm{PL}(\boldsymbol{\theta}) = \prod_{t=1}^{N} f_{Y_t}(y_t|\mathcal{F}_{t-1}). \tag{2.9}$$

Based on equations (2.3) and (2.9), we have the following log partial likelihood for the ZIP autoregression:

$$\begin{aligned} \log \mathrm{PL}(\boldsymbol{\theta}) &= \sum_{t=1}^{N} \log f_{Y_t}(y_t|\mathcal{F}_{t-1}) \\ &= \sum_{t=1}^{N} \log\{\omega_t y_{0,t} + (1-\omega_t)\exp(-\lambda_t)\lambda_t^{y_t}/y_t!\}. \end{aligned}$$

2.2.1 Parameter Estimation

To obtain the MPLE, we need to maximize $\log \mathrm{PL}(\boldsymbol{\theta})$. Equivalently, we must solve the partial score equation $\mathbf{S}_N(\boldsymbol{\theta}) = \mathbf{0}$, where the partial score vector $\mathbf{S}_N(\boldsymbol{\theta})$ is defined as follows:

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

with $\mathbf{C}_{t-1}$ and $\mathbf{v}_t(\boldsymbol{\theta})$ given by

$$\mathbf{C}_{t-1} = \begin{bmatrix} \mathbf{x}_{t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{t-1} \end{bmatrix}$$

and

$$\mathbf{v}_t(\boldsymbol{\theta}) = \begin{bmatrix} v_{1,t}(\boldsymbol{\theta}) \\ v_{2,t}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} y_t - \lambda_t(1-\omega_t y_{0,t}/p_{0,t}) \\ \omega_t(y_{0,t}/p_{0,t} - 1) \end{bmatrix}.$$

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.

Due to the nonlinear nature of the problem, there is no closed-form solution to the partial score equation. Thus, we use iterative algorithms for parameter estimation. We will discuss three different approaches: Newton-Raphson (NR), Fisher Scoring (FS), and Expectation-Maximization (EM) algorithms. Before devising those algorithms, we first present the observed and conditional information matrices. The observed information matrix (i.e., negative Hessian) of the ZIP autoregression is given by

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

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

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

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

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

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

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

It is clear that the elements $\sigma_{11,t}(\boldsymbol{\theta})$ and $\sigma_{22,t}(\boldsymbol{\theta})$ are both positive. Furthermore it can be easily 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 $\omega_t \in (0, 1)$.

To proceed with the NR estimation, we update the estimator through

$$\hat{\boldsymbol{\theta}}^{(j+1)} = \hat{\boldsymbol{\theta}}^{(j)} + \mathbf{H}_N^{-1}(\hat{\boldsymbol{\theta}}^{(j)})\mathbf{S}_N(\hat{\boldsymbol{\theta}}^{(j)}). \tag{2.10}$$

A reasonable starting value $\hat{\boldsymbol{\theta}}^{(0)}$ is needed to start the iteration. We stop the algorithm after some pre-specified convergence criteria are satisfied. In fact, the NR algorithm defined by (2.10) can be viewed as an iteratively reweighted least squares (IRLS) method. To see this, we introduce two design matrices

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_0^\top \\ \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_{N-1}^\top \end{bmatrix} \quad \text{and} \quad \mathbf{Z} = \begin{bmatrix} \mathbf{z}_0^\top \\ \mathbf{z}_1^\top \\ \vdots \\ \mathbf{z}_{N-1}^\top \end{bmatrix}.$$

Furthermore, we let

$$\mathbf{C} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{bmatrix} \quad \text{and} \quad \mathbf{v}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{v}_1(\boldsymbol{\theta}) \\ \mathbf{v}_2(\boldsymbol{\theta}) \end{bmatrix},$$

where $\mathbf{v}_i(\boldsymbol{\theta}) = (v_{i,1}(\boldsymbol{\theta}), \cdots, v_{i,N}(\boldsymbol{\theta}))^\top$ for $i = 1, 2$. Then the partial score vector of the ZIP autoregression can be written as

$$\mathbf{S}_N(\boldsymbol{\theta}) = \sum_{t=1}^{N} \mathbf{C}_{t-1}\mathbf{v}_t(\boldsymbol{\theta}) = \mathbf{C}^\top \mathbf{v}(\boldsymbol{\theta}).$$

In addition, the negative Hessian matrix can be neatly expressed as

$$\mathbf{H}_N(\boldsymbol{\theta}) = \sum_{t=1}^{N} \mathbf{C}_{t-1}\mathbf{D}_t(\boldsymbol{\theta})\mathbf{C}_{t-1}^\top = \mathbf{C}^\top \mathbf{D}(\boldsymbol{\theta})\mathbf{C},$$

where we have

$$\mathbf{D}(\boldsymbol{\theta}) = \begin{bmatrix} \mathrm{Diag}(\mathbf{d}_{11}(\boldsymbol{\theta})) & \mathrm{Diag}(\mathbf{d}_{12}(\boldsymbol{\theta})) \\ \mathrm{Diag}(\mathbf{d}_{12}(\boldsymbol{\theta})) & \mathrm{Diag}(\mathbf{d}_{22}(\boldsymbol{\theta})) \end{bmatrix},$$

with $\mathbf{d}_{ij}(\boldsymbol{\theta}) = (d_{ij,1}(\boldsymbol{\theta}), \cdots, d_{ij,N}(\boldsymbol{\theta}))^\top$ for $i, j = 1, 2$. Note that the matrix $\mathbf{D}(\boldsymbol{\theta})$

contains many zeros and is invertible as long as each individual $\mathbf{D}_t(\boldsymbol{\theta})$ is positive definite. Assuming the inverse of $\mathbf{D}(\boldsymbol{\theta})$ exists, equation (2.10) in the NR algorithm can be rewritten as follows:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}^{(j+1)} &= \mathbf{H}_N^{-1}(\hat{\boldsymbol{\theta}}^{(j)}) \left\{\mathbf{H}_N(\hat{\boldsymbol{\theta}}^{(j)})\hat{\boldsymbol{\theta}}^{(j)} + \mathbf{S}_N(\hat{\boldsymbol{\theta}}^{(j)})\right\} \\
&= \left\{\mathbf{C}^\top \mathbf{D}(\hat{\boldsymbol{\theta}}^{(j)})\mathbf{C}\right\}^{-1} \left\{\mathbf{C}^\top \mathbf{D}(\hat{\boldsymbol{\theta}}^{(j)})\mathbf{C}\hat{\boldsymbol{\theta}}^{(j)} + \mathbf{C}^\top \mathbf{v}(\hat{\boldsymbol{\theta}}^{(j)})\right\} \\
&= \left\{\mathbf{C}^\top \mathbf{D}(\hat{\boldsymbol{\theta}}^{(j)})\mathbf{C}\right\}^{-1} \mathbf{C}^\top \mathbf{D}(\hat{\boldsymbol{\theta}}^{(j)}) \left\{\mathbf{C}\hat{\boldsymbol{\theta}}^{(j)} + \mathbf{D}^{-1}(\hat{\boldsymbol{\theta}}^{(j)})\mathbf{v}(\hat{\boldsymbol{\theta}}^{(j)})\right\} \\
&= \left\{\mathbf{C}^\top \mathbf{D}(\hat{\boldsymbol{\theta}}^{(j)})\mathbf{C}\right\}^{-1} \mathbf{C}^\top \mathbf{D}(\hat{\boldsymbol{\theta}}^{(j)})\mathbf{q}(\hat{\boldsymbol{\theta}}^{(j)}),
\end{aligned} \tag{2.11}$$

where we have $\mathbf{q}(\boldsymbol{\theta}) = \mathbf{C}\boldsymbol{\theta} + \mathbf{D}^{-1}(\boldsymbol{\theta})\mathbf{v}(\boldsymbol{\theta})$. Thus, the updating rule in equation (2.11) is actually a weighted least squares estimation method. By replacing $\mathbf{H}_N(\boldsymbol{\theta})$ by $\mathbf{G}_N(\boldsymbol{\theta})$ in equations (2.10) and (2.11), we obtain the FS algorithm. Based on a parallel development, we show that the FS algorithm may also be viewed as an IRLS method. In practice, the NR and FS algorithms converge very quickly provided that the starting values are close to the MPLE. The performances of these two gradient-based algorithms are generally quite comparable.

Although the NR and FS algorithms can be applied to solve the partial score equation, they are often sensitive to the starting parameter values. To ensure convergence, we take advantage of the mixture structure for the ZIP distribution and estimate the unknown parameters through the EM algorithm (Dempster et al., 1977), an iterative method widely used to fit statistical models involving latent variables. Applying the Bayes theorem to equations (2.5) and (2.6), we have

$$\begin{aligned}
\mathrm{E}(u_t|y_t, \mathcal{F}_{t-1}) &= \Pr(u_t = 1|Y_t = y_t, \mathcal{F}_{t-1}) \\
&= \frac{\Pr(u_t = 1, Y_t = y_t|\mathcal{F}_{t-1})}{\Pr(Y_t = y_t|\mathcal{F}_{t-1})} \\
&= \frac{\Pr(u_t = 1|\mathcal{F}_{t-1})\Pr(Y_t = y_t|u_t = 1, \mathcal{F}_{t-1})}{\Pr(Y_t = y_t|\mathcal{F}_{t-1})} \\
&= \frac{\omega_t y_{0,t}}{\omega_t y_{0,t} + (1-\omega_t)\exp(-\lambda_t)\lambda_t^{y_t}/y_t!},
\end{aligned}$$

which constitutes the basis for the E-step. Up to an additive constant, the complete data log-partial likelihood for $\mathbf{y} = (y_1, \ldots, y_N)^\top$ and $\mathbf{u} = (u_1, \ldots, u_N)^\top$ can be orthogonally decomposed as follows:

$$\begin{aligned}\log \mathrm{PL}^c(\boldsymbol{\theta}) &= \sum_{t=1}^{N}(1-u_t)(y_t \log \lambda_t - \lambda_t) \\ &\quad + \sum_{t=1}^{N}\{u_t \log \omega_t + (1-u_t)\log(1-\omega_t)\}.\end{aligned}$$

The following outline the basic steps of the EM algorithm.

- E-step: Compute the expectation of $\log \mathrm{PL}^c(\boldsymbol{\theta})$ with respect to the conditional distribution of $\mathbf{u}|\mathbf{y}, \boldsymbol{\theta}^{(j)}$. Specifically, we have

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) &= \mathrm{E}\{\log \mathrm{PL}^c(\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\theta}^{(j)}\} \\ &= \sum_{t=1}^{N}(1-\hat{u}_t^{(j)})(y_t \log \lambda_t - \lambda_t) \\ &\quad + \sum_{t=1}^{N}\{\hat{u}_t^{(j)} \log \omega_t + (1-\hat{u}_t^{(j)})\log(1-\omega_t)\},\end{aligned}$$

  where $\hat{u}_t^{(j)}$ denotes the conditional expectation of u_t at j-th iteration.

- M-step: Find $\boldsymbol{\theta}^{(j+1)}$ that maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)})$. Due to the orthogonal partition, we can easily 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. Therefore, the M-step is equivalent to fitting two generalized linear models (i.e., Poisson and logistic regressions).

It is well-known that the EM algorithm tends to slow down when the estimator is very close to the MPLE. To accelerate the convergence speed, we suggest using a hybrid algorithm (EM-NR or EM-FS) by combing the EM algorithm with the NR or FS algorithm. Specifically, we first use the EM algorithm to find appropriate starting values and then switch to the NR or FS algorithm after several EM iterations. This hybrid algorithm has been proved efficient and reliable in many simulated and real examples. Upon convergence, the observed information matrix $\mathbf{H}_N(\boldsymbol{\theta})$ or the conditional information matrix $\mathbf{G}_N(\boldsymbol{\theta})$ can be employed to compute the standard errors of the MPLE. The MPLE together with the standard errors can then be used to form Wald-type tests in the partial likelihood setting.

2.2.2 Asymptotic Theory

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 & Fokianos (2002, Chapter 3).

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 $\mathrm{P}(\sum_{t=1}^{N} \mathbf{C}_{t-1}\mathbf{C}_{t-1}^{\top}$ is positive definite$) = 1$.

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

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

at the true parameter $\boldsymbol{\theta}$.

Fokianos & Kedem (1998) and Kedem & Fokianos (2002, Chapter 3) provide a rigorous treatment of the asymptotic theory for non-stationary categorical time series. Their results are natural extensions of the work by Kaufmann (1987), and their method of proof can be generally applied to any multivariate GLM. For the

ZIP autoregression, we have

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

as $\mathrm{E}(Y_t|\mathcal{F}_{t-1}) = \lambda_t(1-\omega_t)$ and $\mathrm{E}(Y_{0,t}|\mathcal{F}_{t-1}) = p_{0,t}$. Therefore, the partial score process $\{\mathbf{S}_t(\boldsymbol{\theta})\}$, defined by

$$\mathbf{S}_t(\boldsymbol{\theta}) = \sum_{s=1}^{t} \mathbf{C}_{s-1}\mathbf{v}_s(\boldsymbol{\theta}),$$

satisfies the property

$$\begin{aligned}\mathrm{E}\{\mathbf{S}_t(\boldsymbol{\theta})|\mathcal{F}_{t-1}\} &= \mathrm{E}\{\mathbf{S}_{t-1}(\boldsymbol{\theta}) + \mathbf{C}_{t-1}\mathbf{v}_t(\boldsymbol{\theta})|\mathcal{F}_{t-1}\} \\ &= \mathrm{E}\{\mathbf{S}_{t-1}(\boldsymbol{\theta})|\mathcal{F}_{t-1}\} + \mathrm{E}\{\mathbf{C}_{t-1}\mathbf{v}_t(\boldsymbol{\theta})|\mathcal{F}_{t-1}\} = \mathbf{S}_{t-1}(\boldsymbol{\theta}),\end{aligned}$$

indicating that the partial score process $\{\mathbf{S}_t(\boldsymbol{\theta})\}$ is a discrete-time martingale. Moreover the martingale is square integrable since we have

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

These properties together with the martingale central limit theorem (CLT) ensure the consistency and asymptotic normality of the MPLE. More details about the CLT for martingales can be found in the text by Hall & Heyde (1980).

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

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

where $\mathbf{G}(\boldsymbol{\theta})$ is the limiting information matrix per observation such that $\frac{1}{N}\mathbf{G}_N(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{G}(\boldsymbol{\theta})$ as $N \to \infty$.

The proof of Theorem 1 parallels that of Kedem & Fokianos (2002, 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 $F_{Y_t}(m|\mathcal{F}_{t-1};\hat{\boldsymbol{\theta}})$ as stated in the following theorem.

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

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

and

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

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

$$\mathbf{b}_t(\boldsymbol{\theta}) = \begin{bmatrix} \lambda_t(1-\omega_t) \\ -\lambda_t\omega_t(1-\omega_t) \end{bmatrix}$$

and

$$\mathbf{d}_t(\boldsymbol{\theta}) = \begin{bmatrix} (1-\omega_t)\exp(-\lambda_t)\sum_{y_t=0}^{m}(y_t-\lambda_t)\lambda_t^{y_t}/y_t! \\ \omega_t(1-\omega_t)\left\{1-\exp(-\lambda_t)\sum_{y_t=0}^{m}\lambda_t^{y_t}/y_t!\right\} \end{bmatrix}.$$

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

2.2.3 Model Selection Criteria

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 (1974) 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, an analogue of AIC could be defined as follows (Kedem & Fokianos 2002, p. 25):

$$\mathrm{AIC} = -2\log \mathrm{PL}(\hat{\boldsymbol{\theta}}) + 2\dim(\boldsymbol{\theta}), \tag{2.12}$$

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 (Cavanaugh, 1997). 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 (1976) introduced the Takeuchi information criterion (TIC) as an attractive alternative to AIC. In the partial likelihood context, TIC could be defined as

$$\mathrm{TIC} = -2\log \mathrm{PL}(\hat{\boldsymbol{\theta}}) + 2\mathrm{tr}\left\{\mathbf{J}_N(\hat{\boldsymbol{\theta}})\mathbf{I}_N^{-1}(\hat{\boldsymbol{\theta}})\right\}, \tag{2.13}$$

where $\mathbf{I}_N(\boldsymbol{\theta})$ is an information matrix and $\mathbf{J}_N(\boldsymbol{\theta})$ is defined as

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

It is worth noting that the term

$$\mathrm{tr}\left\{\mathbf{J}_N(\hat{\boldsymbol{\theta}})\mathbf{I}_N^{-1}(\hat{\boldsymbol{\theta}})\right\} = \mathrm{tr}\left[\mathbf{I}_N(\hat{\boldsymbol{\theta}})\left\{\mathbf{I}_N^{-1}(\hat{\boldsymbol{\theta}})\mathbf{J}_N(\hat{\boldsymbol{\theta}})\mathbf{I}_N^{-1}(\hat{\boldsymbol{\theta}})\right\}\right]$$

is a measure of discrepancy between the model-based and robust covariance matrices (White, 1982) of the MPLE. For the ZIP autoregression, we have

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

and the information matrix $\mathbf{I}_N(\boldsymbol{\theta})$ can be either $\mathbf{H}_N(\boldsymbol{\theta})$ or $\mathbf{G}_N(\boldsymbol{\theta})$.

It is clear that AIC and TIC share the same goodness-of-fit term. In general, the penalty term $2\mathrm{tr}\left\{\mathbf{J}_N(\hat{\boldsymbol{\theta}})\mathbf{I}_N^{-1}(\hat{\boldsymbol{\theta}})\right\}$ in TIC is very close to the penalty term $2\dim(\boldsymbol{\theta})$ in AIC when the true 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 be much larger than that of AIC. This tendency prevents the selection of a model that is too complicated for the data at hand. It is worth noting that TIC is very similar in structure to the quasi-likelihood information criterion (QIC), which was developed by Pan (2001) for generalized linear model selection where the GLMs are fit using generalized estimating equations (GEEs).

2.3 ZINB Autoregression

In this section, we extend the ZIP autoregression to a more general zero-inflated negative binomial (ZINB) autoregression which can be used to account for simultaneous zero-inflation and overdispersion in many count time series. Conditioning on the filtration $\mathcal{F}_{t-1}$, we assume the response series $\{Y_t\}$ has a ZINB distribution with p.m.f. given by

$$f_{Y_t}(y_t|\mathcal{F}_{t-1}) = \begin{cases} \omega_t + (1-\omega_t)\left(\dfrac{k_t}{k_t+\lambda_t}\right)^{k_t}, & \text{if } y_t = 0, \\ \\ (1-\omega_t)\dfrac{\Gamma(k_t+y_t)}{\Gamma(k_t)y_t!}\left(\dfrac{k_t}{k_t+\lambda_t}\right)^{k_t}\left(\dfrac{\lambda_t}{k_t+\lambda_t}\right)^{y_t}, & \text{if } y_t > 0, \end{cases}$$

where k_t is the dispersion parameter of the baseline NB distribution.

Like the ZIP distribution, the ZINB distribution can be viewed as a two-component mixture of the NB distribution with a degenerate distribution having point mass at zero. Again, we assume there is a dichotomous variable u_t indicating whether the observed count y_t comes from the degenerate distribution ($u_t = 1$) or the ordinary NB distribution ($u_t = 0$). The ZINB distribution can then be equivalently expressed in a hierarchical form where

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

and

$$Y_t|u_t, \mathcal{F}_{t-1} \sim \mathrm{NB}(k_t, (1-u_t)\lambda_t).$$

Based on the hierarchical representation, the conditional mean can be written as

$$\begin{aligned} \mathrm{E}(Y_t|\mathcal{F}_{t-1}) &= \mathrm{E}\{\mathrm{E}(Y_t|u_t, \mathcal{F}_{t-1})\} \\ &= \mathrm{E}\{(1-u_t)\lambda_t|\mathcal{F}_{t-1}\} \\ &= \lambda_t(1-\omega_t), \end{aligned}$$

and the conditional variance can be expressed as

$$\begin{aligned} \mathrm{Var}(Y_t|\mathcal{F}_{t-1}) &= \mathrm{E}\{\mathrm{Var}(Y_t|u_t, \mathcal{F}_{t-1})\} + \mathrm{Var}\{\mathrm{E}(Y_t|u_t, \mathcal{F}_{t-1})\} \\ &= \mathrm{E}\{(1-u_t)\lambda_t + (1-u_t)^2\lambda_t^2/k_t|\mathcal{F}_{t-1}\} + \mathrm{Var}\{(1-u_t)\lambda_t|\mathcal{F}_{t-1}\} \\ &= \lambda_t(1-\omega_t) + \lambda_t^2(1-\omega_t)/k_t + \lambda_t^2\omega_t(1-\omega_t) \\ &= \lambda_t(1-\omega_t)(1+\lambda_t\omega_t+\lambda_t/k_t). \end{aligned}$$

The variance-to-mean ratio or dispersion index

$$\mathrm{Var}(Y_t|\mathcal{F}_{t-1})/\mathrm{E}(Y_t|\mathcal{F}_{t-1}) = 1 + \lambda_t\omega_t + \lambda_t/k_t$$

is greater than or equal to one, and indicates that the overdispersion is attributed to both the zero-inflation parameter ω_t and the dispersion parameter k_t.

In our proposed ZINB autoregression, we model the parameters k_t, λ_t, and ω_t through the following three systematic components:

$$\begin{aligned} \log k_t &= \mathbf{s}_{t-1}^\top \boldsymbol{\alpha}, \\ \log \lambda_t &= \mathbf{x}_{t-1}^\top \boldsymbol{\beta}, \\ \mathrm{logit}(\omega_t) &= \mathbf{z}_{t-1}^\top \boldsymbol{\gamma}, \end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_l)^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ are the regression coefficients. Here $\mathbf{s}_{t-1} = (s_{t-1,1}, \ldots, s_{t-1,l})^\top$, $\mathbf{x}_{t-1} = (x_{t-1,1}, \ldots, x_{t-1,p})^\top$ and $\mathbf{z}_{t-1} =$

$(z_{t-1,1}, \ldots, z_{t-1,q})^\top$ are vectors of past explanatory variables. For convenience, we let $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ denote the $(l+p+q)$-dimensional vector of unknown parameters.

Table 2.1: Observation-driven models for zero-inflated and overdispersed time series.

Zero-Inflation	Overdispersion	Autocorrelation	Model
No	No	No	Poisson Regression
No	No	Yes	Poisson Autoregression
No	Yes	No	NB Regression
No	Yes	Yes	NB Autoregression
Yes	No	No	ZIP Regression
Yes	No	Yes	ZIP Autoregression
Yes	Yes	No	ZINB Regression
Yes	Yes	Yes	ZINB Autoregression

As illustrated in Table 2.1, the ZINB autoregression is very general and includes many important models as special cases. Specifically, ZINB autoregression reduces to

- NB autoregression when $\omega_t = 0$,
- ZIP autoregression when $1/k_t \to 0$,
- Poisson autoregression when $\omega_t = 0$ and $1/k_t \to 0$.

In the absence of autocorrelation, the Poisson, NB, ZIP, and ZINB autoregressions simply reduce to the four ordinary regression models for independent data.

We now consider partial likelihood inference for the most general ZINB autoregression. To simplify the notation, we let $p_t = k_t/(k_t + \lambda_t)$. Then the log-partial likelihood of the ZINB autoregression is given by

$$\begin{aligned}\log \mathrm{PL}(\boldsymbol{\theta}) &= \sum_{y_t=0} \log\{\omega_t + (1-\omega_t)p_t^{k_t}\} \\ &\quad + \sum_{y_t>0} \{\log(1-\omega_t) + \log\Gamma(k_t+y_t) - \log\Gamma(k_t) \\ &\qquad\qquad - \log(y_t!) + k_t \log p_t + y_t \log(1-p_t)\}\,.\end{aligned}$$

It can be easily verified that

$$\begin{aligned}\partial p_t/\partial k_t &= p_t(1-p_t)/k_t, \\ \partial p_t/\partial \lambda_t &= -p_t^2/k_t, \\ \partial p_t^{k_t}/\partial k_t &= p_t^{k_t}(\log p_t + 1 - p_t), \\ \partial p_t^{k_t}/\partial \lambda_t &= -p_t^{k_t+1}.\end{aligned}$$

Then the partial score vector of the ZINB autoregression is given by

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

where we have

$$\mathbf{C}_{t-1} = \begin{bmatrix} \mathbf{s}_{t-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{z}_{t-1} \end{bmatrix}, \quad \mathbf{v}_t(\boldsymbol{\theta}) = \begin{bmatrix} v_{1,t}(\boldsymbol{\theta}) \\ v_{2,t}(\boldsymbol{\theta}) \\ v_{3,t}(\boldsymbol{\theta}) \end{bmatrix}.$$

Here the elements of $\mathbf{v}_t(\boldsymbol{\theta})$ are defined as follows:

$$\begin{aligned}v_{1,t}(\boldsymbol{\theta}) &= k_t(1-\omega_t y_{0,t}/p_{0,t})(\log p_t + 1 - p_t) \\ &\quad + k_t(1-y_{0,t})\{\Psi_0(k_t+y_t) - \Psi_0(k_t)\} - y_t p_t, \\ v_{2,t}(\boldsymbol{\theta}) &= p_t y_t - k_t(1-p_t)(1-\omega_t y_{0,t}/p_{0,t}), \\ v_{3,t}(\boldsymbol{\theta}) &= \omega_t(y_{0,t}/p_{0,t} - 1),\end{aligned}$$

where $\Psi_0(\cdot)$ is the digamma function, and $p_{0,t} = \omega_t + (1-\omega_t)p_t^{k_t}$ is the p.m.f. of

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

$$\begin{aligned}
\mathrm{E}\{v_{1,t}(\boldsymbol{\theta})|\mathcal{F}_{t-1}\} &= k_t(1-\omega_t p_{0,t}/p_{0,t})(\log p_t + 1 - p_t) \\
&\quad + k_t \mathrm{E}[(1-y_{0,t})\{\Psi_0(k_t+y_t) - \Psi_0(k_t)\}|\mathcal{F}_{t-1}] - (1-\omega_t)\lambda_t p_t \\
&= k_t(1-\omega_t)\log p_t + k_t \mathrm{E}[(1-y_{0,t})\{\Psi_0(k_t+y_t) - \Psi_0(k_t)\}|\mathcal{F}_{t-1}] \\
&\quad + k_t(1-\omega_t)\frac{\lambda_t}{k_t+\lambda_t} - (1-\omega_t)\lambda_t \frac{k_t}{k_t+\lambda_t} \\
&= k_t(1-\omega_t)\log p_t + k_t \mathrm{E}[(1-y_{0,t})\{\Psi_0(k_t+y_t) - \Psi_0(k_t)\}|\mathcal{F}_{t-1}], \\
\mathrm{E}\{v_{2,t}(\boldsymbol{\theta})|\mathcal{F}_{t-1}\} &= p_t(1-\omega_t)\lambda_t - k_t(1-p_t)(1-\omega_t p_{0,t}/p_{0,t}) \\
&= \frac{k_t}{k_t+\lambda_t}(1-\omega_t)\lambda_t - k_t\frac{\lambda_t}{k_t+\lambda_t}(1-\omega_t) = 0, \\
\mathrm{E}\{v_{3,t}(\boldsymbol{\theta})|\mathcal{F}_{t-1}\} &= \omega_t(p_{0,t}/p_{0,t} - 1) = 0.
\end{aligned}$$

Thus, for the partial score process of the ZINB autoregression to be a zero-mean martingale, we still need to show $\mathrm{E}\{v_{1,t}(\boldsymbol{\theta})|\mathcal{F}_{t-1}\} = 0$, or equivalently

$$\begin{aligned}
& \mathrm{E}[(1-y_{0,t})\{\Psi_0(k_t) - \Psi_0(k_t+y_t)\}|\mathcal{F}_{t-1}] \\
=\; & -\sum_{j=0}^{\infty} \Pr(Y_t > j|\mathcal{F}_{j-1})/(k_t+j) \\
=\; & (1-\omega_t)\log p_t. \qquad (2.14)
\end{aligned}$$

We have verified equation (2.14) for several important special cases when $k_t = 1, 2, 3$. In addition, we have validated the general conjecture replying on computers. However, a complete and rigorous proof of the conjecture has yet to be formulated.

The observed information matrix of the ZINB autoregression is given by

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

where $\mathbf{D}_t(\boldsymbol{\theta})$ is a symmetric 3×3 matrix with elements defined as follows:

$$
\begin{aligned}
d_{11,t}(\boldsymbol{\theta}) &= k_t(1-p_t)\{y_t p_t/k_t - (1-p_t)(1-\omega_t y_{0,t}/p_{0,t})\} \\
&\quad - k_t(1-y_{0,t})\{\Psi_0(k_t+y_t) - \Psi_0(k_t) + k_t\Psi_1(k_t+y_t) - k_t\Psi_1(k_t)\} \\
&\quad - k_t(\log p_t + 1 - p_t)\,\{1 - \omega_t y_{0,t}/p_{0,t} \\
&\qquad\qquad + k_t\omega_t y_{0,t}(\log p_t + 1 - p_t)(p_{0,t}-\omega_t)/p_{0,t}^2\}\,, \\
d_{12,t}(\boldsymbol{\theta}) &= k_t(1-p_t)\,\big\{k_t\omega_t y_{0,t}(\log p_t + 1 - p_t)(p_{0,t}-\omega_t)/p_{0,t}^2 \\
&\qquad\qquad + (1-p_t)(1-\omega_t y_{0,t}/p_{0,t}) - y_t p_t/k_t\}\,, \\
d_{13,t}(\boldsymbol{\theta}) &= k_t\omega_t y_{0,t}(\log p_t + 1 - p_t)(p_{0,t}-\omega_t)/p_{0,t}^2, \\
d_{22,t}(\boldsymbol{\theta}) &= k_t(1-p_t)\,\{p_t y_t/k_t + p_t(1-\omega_t y_{0,t}/p_{0,t}) \\
&\qquad\qquad - k_t\omega_t y_{0,t}(1-p_t)(p_{0,t}-\omega_t)/p_{0,t}^2\}\,, \\
d_{23,t}(\boldsymbol{\theta}) &= -k_t\omega_t(1-\omega_t)(1-p_t)p_t^{k_t} y_{0,t}/p_{0,t}^2, \\
d_{33,t}(\boldsymbol{\theta}) &= \omega_t(1-\omega_t)(1-p_t^{k_t} y_{0,t}/p_{0,t}^2).
\end{aligned}
$$

Here $\Psi_1(\cdot)$ is the trigamma function. The conditional information matrix of the ZINB autoregression is given by

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

where $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is a symmetric 3×3 matrix with elements defined as follows:

$$
\begin{aligned}
\sigma_{11,t}(\boldsymbol{\theta}) &= k_t\{k_t c_t - (1-\omega_t)(1-p_t)\} - \omega_t(1-\omega_t)k_t^2(\log p_t + 1 - p_t)^2 p_t^{k_t}/p_{0,t}, \\
\sigma_{12,t}(\boldsymbol{\theta}) &= \omega_t(1-\omega_t)k_t^2(1-p_t)(\log p_t + 1 - p_t)p_t^{k_t}/p_{0,t}, \\
\sigma_{13,t}(\boldsymbol{\theta}) &= \omega_t(1-\omega_t)k_t(\log p_t + 1 - p_t)p_t^{k_t}/p_{0,t}, \\
\sigma_{22,t}(\boldsymbol{\theta}) &= (1-\omega_t)k_t(1-p_t)[p_t^{k_t} + \omega_t\{1 - p_t^{k_t} - k_t(1-p_t)p_t^{k_t}\}]/p_{0,t}, \\
\sigma_{23,t}(\boldsymbol{\theta}) &= -\omega_t(1-\omega_t)k_t(1-p_t)p_t^{k_t}/p_{0,t}, \\
\sigma_{33,t}(\boldsymbol{\theta}) &= \omega_t^2(1-\omega_t)(1-p_t^{k_t})/p_{0,t}.
\end{aligned}
$$

The term c_t that appears in $\sigma_{11,t}(\boldsymbol{\theta})$ is defined as follows:

$$\begin{aligned} c_t &= \mathrm{E}\{(1-y_{0,t})[\Psi_1(k_t) - \Psi_1(k_t+y_t)|\mathcal{F}_{j-1}]\} \\ &= \sum_{j=0}^{\infty} \Pr(Y_t > j|\mathcal{F}_{j-1})/(k_t+j)^2. \end{aligned}$$

Taking the derivative with respect to k_t on both sides of (2.14), we should expect c_t to be equal to $(1-\omega_t)(1-p_t)/k_t$, provided that the differentiation and infinite summation can be exchanged. However, this exchangeability assumption does not always hold for (2.14). Fortunately, we can easily approximate c_t by

$$\tilde{c}_t = \sum_{j=0}^{M} \Pr(Y_t > j|\mathcal{F}_{j-1})/(k_t+j)^2,$$

where M a reasonably large integer (e.g., 100).

The results pertaining to parameter estimation, asymptotic theory, and model selection for ZINB autoregression parallel those for ZIP autoregression. For the sake of brevity, these results are not reproduced here.

2.4 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 ZIP autoregression with

$$\begin{aligned} \eta_t &= \beta_0 + \beta_1 I_{(y_{t-1}>0)} + \sigma v_t, \\ \xi_t &= \gamma_0 + \gamma_1 I_{(y_{t-1}>0)}. \end{aligned}$$

Here, v_t is an unobservable realization from the standard normal distribution, included in η_t to optionally induce extra overdispersion in the data. Other functions of y_{t-1} (e.g., $\log(y_{t-1}+1)$) can also be employed in η_t and ξ_t to account for autocorrelation. We assume $\boldsymbol{\theta} = (1.2, 0.6, 0.4, -0.8)^\top$ is the true parameter vector.

Table 2.2: Finite sample results of the MPLE based on 1,000 replications simulated independently from the true model 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

The first set of simulation results are compiled under correct model specification (i.e., $\sigma = 0$). Table 2.2 summarizes the finite sample results of the MPLE based on three different sample sizes (N = 100, 200, 500). The observed information matrix $\mathbf{H}_N(\boldsymbol{\theta})$ is employed for the computation of the large-sample covariance matrix which is subsequently used to calculate the model-based asymptotic standard errors (ASEs). From Table 2.2, we can see the absolute bias decreases as the sample size increases. The means of the 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. It is worth noting that the absoluate bias, ASE, and ESD tend to be larger for the logistic component (2.8) than for the log-linear component (2.7). 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 (Yau et al., 2004). The Q-Q plots in Figure 2.1 confirm the asymptotic normality of the MPLE.

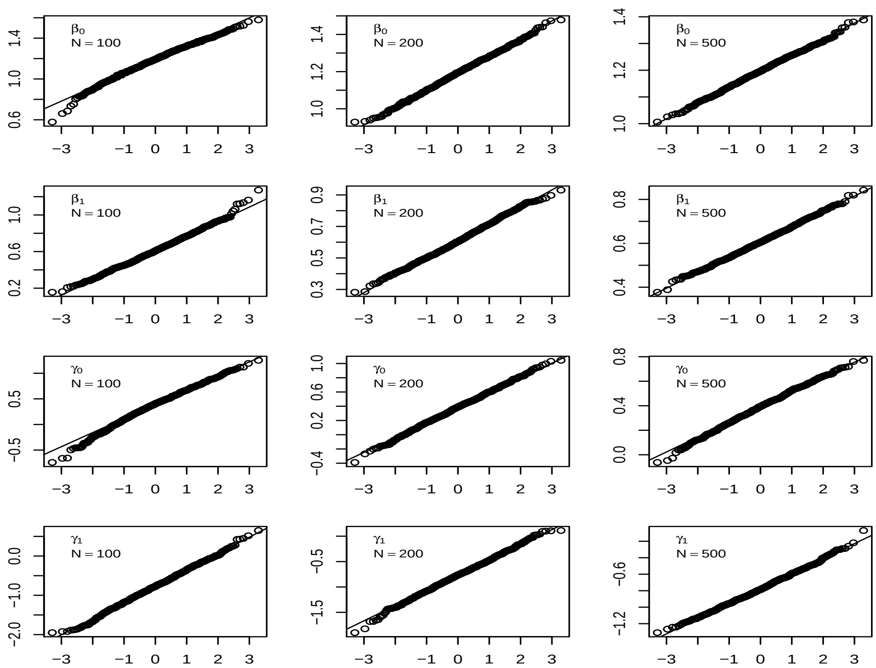

Figure 2.1: Q-Q plots for the estimated parameters based on 1,000 replications.

We next investigate the variable selection problem. We assume the data is generated from the true model 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:

$$\eta_t = \beta_0 + \sum_{i=1}^{k_1} \beta_i I_{(y_{t-i}>0)}$$

and

$$\xi_t = \gamma_0 + \sum_{i=1}^{k_2} \gamma_i I_{(y_{t-i}>0)}$$

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

Table 2.3 summarizes the model selection results based on 1,000 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 2.3). Thus, TIC clearly outperforms AIC in the presence of model misspecification, as it tends to offer greater protection against unwarranted complexity. Under correct distributional specification (i.e., $\sigma = 0$), the performances of AIC and TIC are quite comparable (results not shown).

Table 2.3: Variable selection results of TIC and AIC (in parentheses) based on 1,000 replications simulated independently from the true model 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)

2.5 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. Syphilis is a localized infection and it is relatively common in the South (Table 2.4). For states where syphilis is less common, the temporal surveillance for the disease is complicated due to the small counts collected.

Table 2.4: Top ten states ranked by syphilis rate (per 100,000) in 2009.

Rank	State	Rate
1	Louisiana	16.8
2	Georgia	9.8
3	Arkansas	9.6
4	Alabama	8.9
5	Mississippi	8.1
6	Texas	6.8
7	Tennessee	6.5
8	North Carolina	6.3
9	New York	6.1
10	Illinois	5.8

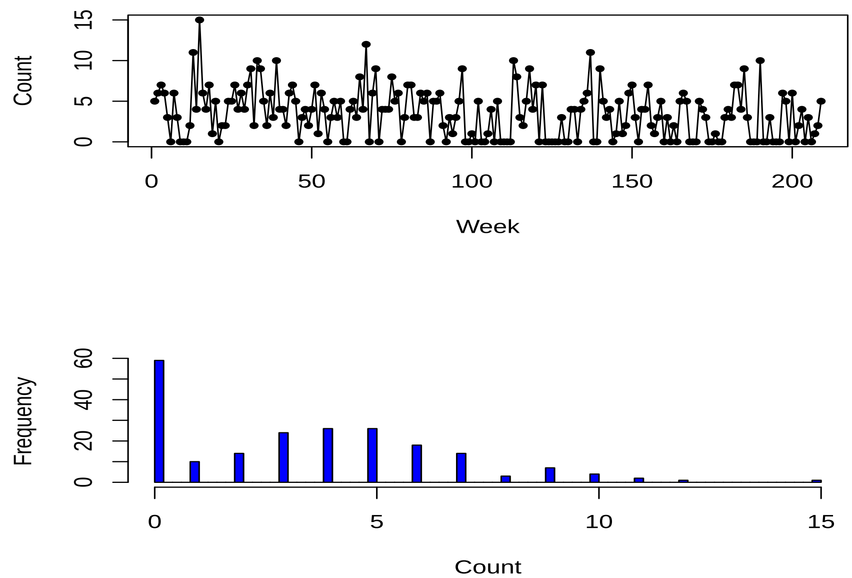

Figure 2.2: Time series plot (top) and histogram (bottom) for weekly syphilis counts in Maryland from 2007 to 2010 (http://www.cdc.gov/mmwr/).

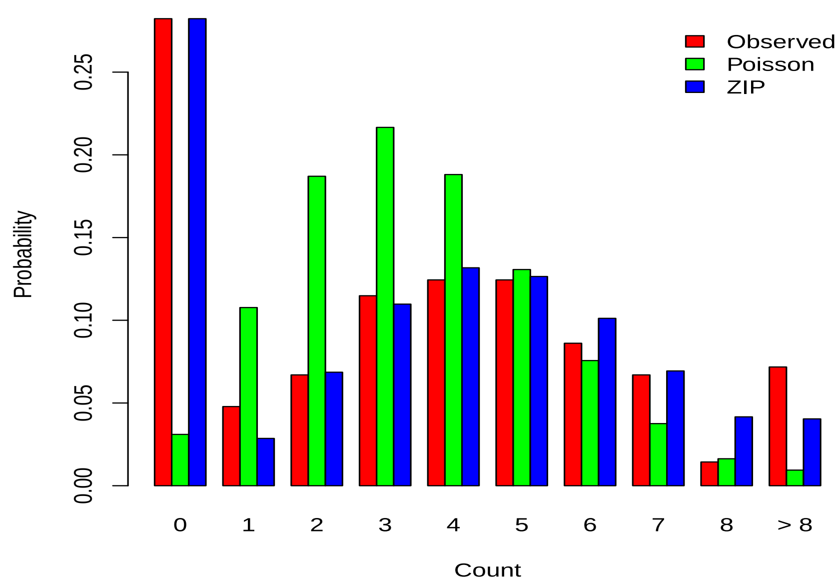

Figure 2.3: Goodness-of-fit for the syphilis data from Maryland.

Our data consists of the weekly number of syphilis cases reported in Maryland. The series is extracted from the CDC's Morbidity and Mortality Weekly Report (MMWR). Figure 2.2 displays the time series plot and histogram of the syphilis incidence for the period of 2007-2010, during which a large number of zeros 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.

We first select an appropriate parametric distribution for the syphilis data. Specifically, we compare the empirical distribution of the data to a fitted ZIP distribution and its Poisson counterpart (see Figure 2.3). The fitted probabilities for the ZIP distribution closely follow the observed probabilities for the syphilis data, while the fitted probabilities for the Poisson distribution severely underestimate or overestimate the empirical probabilities, especially when the count is equal to zero. To account for the serial correlation and the preponderance of zeros, we next consider the following autoregressive models based on the ZIP distribution:

$$\eta_t = \beta_0 + \sum_{i=1}^{k_1} \beta_i I_{(y_{t-i}>0)} + \beta_{k_1+1} x_t$$

and

$$\xi_t = \gamma_0 + \sum_{i=1}^{k_2} \gamma_i I_{(y_{t-i}>0)} + \gamma_{k_2+1} x_t,$$

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.

Figure 2.4 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:

$$\eta_t = \beta_0 + \beta_1 I_{(y_{t-1}>0)} + \beta_2 x_t$$

and

$$\xi_t = \gamma_0 + \gamma_1 x_t.$$

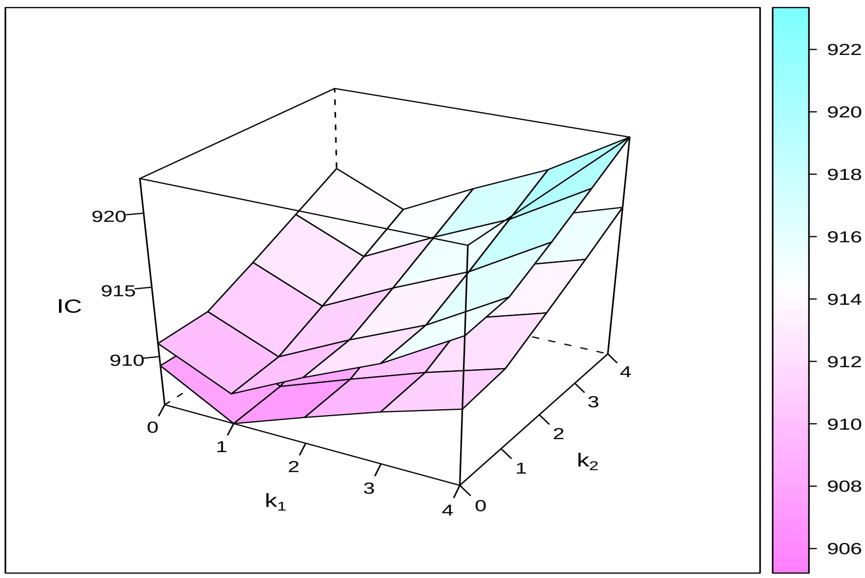

Figure 2.4: AIC (bottom) and TIC (top) values for the 25 ZIP candidate models fit to the 2007-2010 syphilis data, with $k_1, k_2 = 0, \ldots, 4$. The penalties of TIC are uniformly larger than those of AIC. The model with $(k_1, k_2) = (1, 0)$ is favored by both AIC and TIC.

Table 2.5 summarizes the output for final ZIP model and its Poisson counterpart. Here the p-values for the regression coefficients are based on the Wald-type tests. Both models suggest an AR(1) component for the log-linear part, but the downward trend is found to be significant only in the Poisson autoregression. 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 autoregression. Thus, the significant downward trend ($p < 0.0001$) obtained with this model should not be trusted. On the other hand, the ZIP autoregression also detects a downward but nonsignificant trend in the log-linear systematic component ($p = 0.1299$). According to this model, we observe more zeros in the latter period of the series (Figure 2.2) due to the significant increase of the zero-inflation parameter in the logistic systematic component ($p = 0.0022$).

Table 2.5: Final ZIP autoregression and its Poisson counterpart for the Maryland syphilis data (2007 to 2010).

	ZIP Model (TIC = 920.8, AIC = 918.8)			Poisson Model (TIC = 1130.3, AIC = 1120.9)		
$\boldsymbol{\theta}$	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 practice, public health officials are often interested in forecasting future disease trends. This task can be easily accomplished based on the final ZIP model. As an illustration, we consider a one-step-ahead prediction where the model parameters are sequentially updated once a new observation accrues. 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. Figure 2.5 features the plot of predictive probabilities based on a cutoff of 6 for the first 36 weeks of 2011 in Maryland. The ZIP autoregression closely tracks the average outbreak probability, while the Poisson autoregression, not surprisingly, underestimates the average probability. 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.

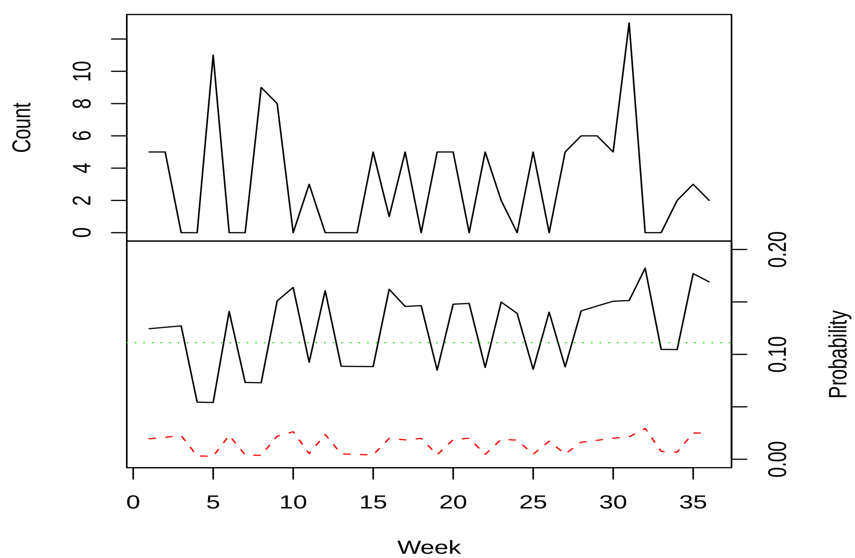

Figure 2.5: Time series plot (top) and one-step-ahead predictive probabilities (bottom) for the first 36 weeks of 2011. The ZIP autoregression (black solid line) closely tracks the average outbreak probability (green dotted line), while the Poisson autoregression (red dashed line) underestimates the average probability.

CHAPTER 3
PARAMETER-DRIVEN MODELS

In this chapter, we focus on parameter-driven models or dynamic models for zero-inflated time series. We first review the state-space model (Kalman, 1960; Kalman & Bucy, 1961) for normally distributed data. In time series literature, such a model is often called the dynamic linear model (DLM) and is a natural extension of the traditional linear model. After reviewing the DLM, we introduce a class of dynamic models for count time series with a high frequency of zeros. To estimate the model parameters, we devise a Monte Carlo EM (MCEM) algorithm, where particle filtering and smoothing methods (Gordon et al., 1993; Godsill, et al., 2004) are employed to approximate high-dimensional integrals. Both simulated and real examples are presented to illustrate the proposed dynamic models.

3.1 Dynamic Linear Model

Although the DLM was originally introduced as a method for aerospace-related research, it has been subsequently applied to a variety of other disciplines such as economics, ecology, and medicine. The DLM is a very general and flexible modeling framework. Many popular time series models, including the ARIMA models, belong to the class of DLMs and thus can be fit through a state-space approach. Here, we only attempt to give a brief overview of the DLM. A rigorous treatment of the topic can be found in the texts by Durbin & Koopman (2001) and Shumway & Stoffer (2006, Chapter 6).

Basically, the DLM assumes the existence of an unobservable state $\mathbf{s}_t$ that evolves as a first-order vector autoregression

$$\mathbf{s}_t|\mathbf{s}_{t-1} \sim \mathcal{N}_p(\mathbf{\Phi}\mathbf{s}_{t-1} + \mathbf{\Upsilon}\mathbf{x}_t, \mathbf{Q}), \tag{3.1}$$

where $\mathbf{x}_t$ is an $r \times 1$ vector of inputs. Here $\mathbf{\Phi}$ is $p \times p$, $\mathbf{\Upsilon}$ is $p \times r$, and $\mathbf{Q}$ is $p \times p$.

In the DLM, we assume the initial state $\mathbf{s}_0$ is normally distributed with mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. Conditioning on the current state $\mathbf{s}_t$, the observation $\mathbf{y}_t$ is assumed to be independently distributed as follows:

$$\mathbf{y}_t|\mathbf{s}_t \sim \mathcal{N}_q(\mathbf{A}_t\mathbf{s}_t + \boldsymbol{\Gamma}\mathbf{x}_t, \mathbf{R}), \tag{3.2}$$

where $\mathbf{A}_t$ is $q \times p$, $\boldsymbol{\Gamma}$ is $q \times r$, and $\mathbf{R}$ is $q \times q$. Here we let $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \boldsymbol{\Phi}, \boldsymbol{\Upsilon}, \mathbf{Q}, \boldsymbol{\Gamma}, \mathbf{R}\}$ denote the set of unknown parameters. The time-varying matrix $\mathbf{A}_t$ in (3.2) is assumed to be known. Figure 3.1 illustrates the state evolution and data generation in the DLM that is defined by (3.1) and (3.2).

$$\begin{array}{ccccccccc}
\mathbf{s}_0 & \longrightarrow & \mathbf{s}_1 & \longrightarrow & \mathbf{s}_2 & \longrightarrow & \cdots & \longrightarrow & \mathbf{s}_n \\
 & & \downarrow & & \downarrow & & & & \downarrow \\
 & & \mathbf{y}_1 & & \mathbf{y}_2 & & & & \mathbf{y}_n
\end{array}$$

Figure 3.1: Graphical illustration of the state evolution and data generation in the dynamic linear model.

Our goal here is to find the maximum likelihood estimator (MLE) of $\boldsymbol{\Theta}$ through the observable data $\mathbf{y}_{1:n}$. Since the marginal likelihood of $\mathbf{y}_{1:n}$ cannot be expressed analytically, Shumway and Stoffer (1982) introduced an EM algorithm to estimate $\boldsymbol{\Theta}$ based on the Kalman filtering and smoothing techniques. A comprehensive discussion of the Kalman methods can be found in the text by Shumway and Stoffer (2006, Chapter 6). Here we extend the EM algorithm by Shumway and Stoffer (1982) to the general setting where exogenous variables $\mathbf{x}_t$ are included in the state and observation equations. Assuming we were able to observe the latent state $\mathbf{s}_{0:n}$, the complete-data likelihood (i.e., the joint density of $\mathbf{s}_{0:n}$ and $\mathbf{y}_{1:n}$) can be orthogonally decomposed as follows:

$$\begin{aligned} L_c(\boldsymbol{\Theta}) &= f(\mathbf{s}_{0:n}, \mathbf{y}_{1:n}) \\ &= f(\mathbf{s}_{0:n}) f(\mathbf{y}_{1:n}|\mathbf{s}_{0:n}) \\ &= f(\mathbf{s}_0) \prod_{t=1}^{n} f(\mathbf{s}_t|\mathbf{s}_{t-1}) \prod_{t=1}^{n} f(\mathbf{y}_t|\mathbf{s}_t). \end{aligned}$$

Up to an additive constant, we have

$$\begin{aligned} -2\log L_c(\boldsymbol{\Theta}) &= \log|\boldsymbol{\Sigma}_0| + (\mathbf{s}_0 - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\mathbf{s}_0 - \boldsymbol{\mu}_0) \\ &\quad + n\log|\mathbf{Q}| + \sum_{t=1}^{n}(\mathbf{s}_t - \boldsymbol{\Phi}\mathbf{s}_{t-1} - \boldsymbol{\Upsilon}\mathbf{x}_t)^\top \mathbf{Q}^{-1}(\mathbf{s}_t - \boldsymbol{\Phi}\mathbf{s}_{t-1} - \boldsymbol{\Upsilon}\mathbf{x}_t) \\ &\quad + n\log|\mathbf{R}| + \sum_{t=1}^{n}(\mathbf{y}_t - \mathbf{A}_t\mathbf{s}_t - \boldsymbol{\Gamma}\mathbf{x}_t)^\top \mathbf{R}^{-1}(\mathbf{y}_t - \mathbf{A}_t\mathbf{s}_t - \boldsymbol{\Gamma}\mathbf{x}_t). \end{aligned}$$

To implement the EM algorithm, we need to calculate the following expectations given the observed data

$$\begin{aligned} \mathbf{s}_t^n &= \mathrm{E}(\mathbf{s}_t|\mathbf{y}_{0:n}), \\ \mathbf{P}_t^n &= \mathrm{E}\{(\mathbf{s}_t - \mathbf{s}_t^n)(\mathbf{s}_t - \mathbf{s}_t^n)^\top|\mathbf{y}_{0:n}\}, \\ \mathbf{P}_{t,t-1}^n &= \mathrm{E}\{(\mathbf{s}_t - \mathbf{s}_t^n)(\mathbf{s}_{t-1} - \mathbf{s}_{t-1}^n)^\top|\mathbf{y}_{0:n}\}. \end{aligned}$$

The recursive formulas for computing the above expectations are available in the text by Shumway and Stoffer (2006, pp. 330-339). In the following EM algorithm, $\mathbf{s}_t^n$, $\mathbf{P}_t^n$, and $\mathbf{P}_{t,t-1}^n$ are all evaluated at $\boldsymbol{\Theta}^{(j)}$.

- In the E-step, we compute

$$\begin{aligned} Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(j)}) &= \mathrm{E}\{-2\log L_c(\boldsymbol{\Theta})|\mathbf{y}_{0:n}, \boldsymbol{\Theta}^{(j)}\} \\ &= \log|\boldsymbol{\Sigma}_0| + \mathrm{tr}\left[\boldsymbol{\Sigma}_0^{-1}\left\{\mathbf{P}_0^n + (\mathbf{s}_0^n - \boldsymbol{\mu}_0)(\mathbf{s}_0^n - \boldsymbol{\mu}_0)^\top\right\}\right] \\ &\quad + n\log|\mathbf{Q}| + \mathrm{tr}\left\{\mathbf{Q}^{-1}(\mathbf{S}_{11} - \mathbf{S}_{10}\boldsymbol{\Psi}^\top - \boldsymbol{\Psi}\mathbf{S}_{10}^\top + \boldsymbol{\Psi}\mathbf{S}_{00}\boldsymbol{\Psi}^\top)\right\} \\ &\quad + n\log|\mathbf{R}| + \mathrm{tr}\left\{\mathbf{R}^{-1}(\mathbf{T}_{11} - \mathbf{T}_{10}\boldsymbol{\Gamma}^\top - \boldsymbol{\Gamma}\mathbf{T}_{10}^\top + \boldsymbol{\Gamma}\mathbf{T}_{00}\boldsymbol{\Gamma}^\top)\right\}, \end{aligned}$$

where $\mathbf{\Psi} = [\mathbf{\Phi} \ \ \mathbf{\Upsilon}]$ is a $p \times (p+r)$ matrix. In addition, we have

$$\begin{aligned}
\mathbf{S}_{11} &= \sum_{t=1}^{n} (\mathbf{s}_t^n \mathbf{s}_t^{n\top} + \mathbf{P}_t^n), \\
\mathbf{S}_{10} &= \sum_{t=1}^{n} \left[\begin{array}{cc} \mathbf{s}_t^n \mathbf{s}_{t-1}^{n\ \top} + \mathbf{P}_{t,t-1}^n & \mathbf{s}_t^n \mathbf{x}_t^\top \end{array} \right], \\
\mathbf{S}_{00} &= \sum_{t=1}^{n} \left[\begin{array}{cc} \mathbf{s}_{t-1}^n \mathbf{s}_{t-1}^{n\ \top} + \mathbf{P}_{t-1}^n & \mathbf{s}_{t-1}^n \mathbf{x}_t^\top \\ \mathbf{x}_t \mathbf{s}_{t-1}^{n\ \top} & \mathbf{x}_t \mathbf{x}_t^\top \end{array} \right], \\
\mathbf{T}_{11} &= \sum_{t=1}^{n} \{ (\mathbf{y}_t - \mathbf{A}_t \mathbf{s}_t^n)(\mathbf{y}_t - \mathbf{A}_t \mathbf{s}_t^n)^\top + \mathbf{A}_t \mathbf{P}_t^n \mathbf{A}_t^\top \}, \\
\mathbf{T}_{10} &= \sum_{t=1}^{n} (\mathbf{y}_t - \mathbf{A}_t \mathbf{s}_t^n) \mathbf{x}_t^\top, \\
\mathbf{T}_{00} &= \sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t^\top.
\end{aligned}$$

- In the M-step, we update the parameters as follows:

$$\begin{aligned}
\boldsymbol{\mu}_0^{(j+1)} &= \mathbf{s}_0^n, \\
\mathbf{\Sigma}_0^{(j+1)} &= \mathbf{P}_0^n, \\
\mathbf{\Psi}^{(j+1)} &= \mathbf{S}_{10} \mathbf{S}_{00}^{-1}, \\
\mathbf{Q}^{(j+1)} &= n^{-1} (\mathbf{S}_{11} - \mathbf{S}_{10} \mathbf{S}_{00}^{-1} \mathbf{S}_{10}^\top), \\
\mathbf{\Gamma}^{(j+1)} &= \mathbf{T}_{10} \mathbf{T}_{00}^{-1}, \\
\mathbf{R}^{(j+1)} &= n^{-1} (\mathbf{T}_{11} - \mathbf{T}_{10} \mathbf{T}_{00}^{-1} \mathbf{T}_{10}^\top).
\end{aligned}$$

We can obtain the MLE by iteratively applying the preceding E-step and M-step. Harvey (1989, pp. 140-142) provides a formula to compute the observed information matrix. Cavanaugh & Shumway (1996) present a general and recursive algorithm to calculate the expected information matrix (i.e., Fisher information matrix). In practice, either the observed or expected information matrix can be used to compute the standard errors of the MLE.

3.2 Dynamic ZIP Model

The DLM is a very useful tool that has been widely used in many different disciplines. However, it cannot be applied to analyze time series with low counts, especially for count series with a large proportion of zeros. In this section, we propose a dynamic ZIP model to accommodate zero-inflation in count time series. Specifically, we assume there is a stationary AR(p) process $\{z_t\}$ such that

$$z_t = \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} + \epsilon_t,$$

where ϵ_t is a white noise process with mean 0 and variance σ^2. Conditioning on the current state z_t, we assume that the observation y_t has a ZIP distribution with p.m.f.

$$f_{Y_t}(y_t|z_t; \lambda_t, \omega) = \begin{cases} \omega + (1-\omega)\exp(-\lambda_t), & \text{if } y_t = 0, \\ \\ (1-\omega)\exp(-\lambda_t)\lambda_t^{y_t}/y_t!, & \text{if } y_t > 0. \end{cases}$$

We use the following log-linear model to characterize the intensity parameter λ_t:

$$\log \lambda_t = \log w_t + \mathbf{x}_t^\top \boldsymbol{\beta} + z_t,$$

where $\mathbf{x}_t$ is a set of explanatory variables and $\boldsymbol{\beta}$ is the vector of regression coefficients. Here, $\log w_t$ is referred to as the offset variable. Let $\boldsymbol{\theta} = (\omega, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \sigma)^\top$ denote the vector of unknown parameters. For simplicity, the zero-inflation parameter ω is treated as constant, although it can also be described by a separate logistic or probit model.

The preceding dynamic ZIP model is very general. It includes as special cases

- the ZIP mixed autoregression (Yau et al., 2004) when $p = 1$;
- the Poisson state-space model (Chan & Ledolter, 1995) when $p = 1$ and $\omega = 0$.

3.2.1 State-Space Representation

The dynamic ZIP model can be written in the following state-space form:

$$\mathbf{s}_t|\mathbf{s}_{t-1} \sim \mathcal{N}_p(\mathbf{\Phi}\mathbf{s}_{t-1}, \mathbf{\Sigma}), \tag{3.3}$$

$$u_t \sim \text{Bernoulli}(\omega), \tag{3.4}$$

$$y_t|\mathbf{s}_t, u_t \sim \text{Poisson}((1-u_t)\lambda_t), \tag{3.5}$$

where $\mathbf{s}_t = (z_t, \ldots, z_{t-p+1})^\top$ is a p-dimensional state vector and u_t is an unobservable membership indicator. Similar to the DLM, the initial state $\mathbf{s}_0$ is assumed to be normally distributed with mean $\boldsymbol{\mu}_0$ and covariance matrix $\mathbf{\Sigma}_0$. Here $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are $p \times p$ matrices defined as follows:

$$\mathbf{\Phi} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \qquad \mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{3.6}$$

Note that the covariance matrix $\mathbf{\Sigma}$ in (3.6) is not positive definite. This is legitimate in the state-space modeling approach. Figure 3.2 illustrates the dynamic ZIP model that is defined by (3.3)-(3.5).

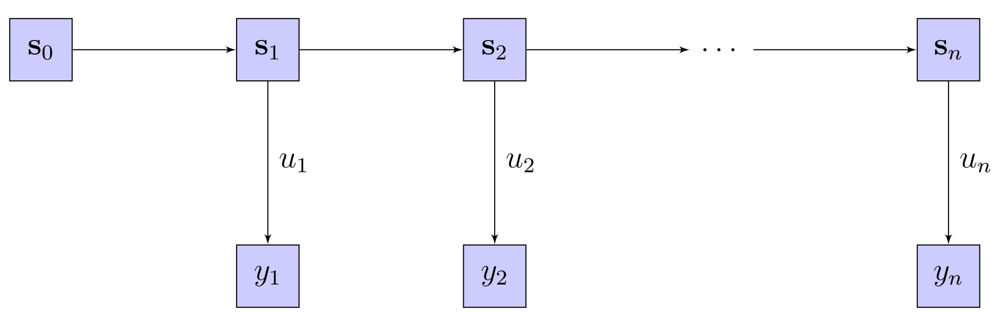

Figure 3.2: Graphical illustration of the state evolution and data generation in the dynamic ZIP model.

Based on the state-space representation, we orthogonally decompose the complete-data likelihood (i.e., the joint density of $\mathbf{s}_{0:n}$, $u_{1:n}$, and $y_{1:n}$) as follows:

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

Since the marginal likelihood of $y_{1:n}$ cannot be expressed analytically, we employ an EM algorithm to estimate $\boldsymbol{\theta}$. Up to an additive constant, the complete-data log-likelihood is given by

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

To implement the EM algorithm, we need to compute the expectation of $l_c(\boldsymbol{\theta})$ given the observed data $y_{1:n}$. Unlike the DLM, there is no analytical form for the conditional expectation due to the non-normality of the data. To approximate the conditional expectation in the E-step, we resort to particle filtering and smoothing techniques (Gordon, 1993; Godsill et al., 2004), which are Monte Carlo extensions of the well-known Kalman methods.

3.2.2 Particle Methods

Particle filtering and particle smoothing belong to the class of sequential Monte Carlo (SMC) methods. The basic idea behind particle methods is to approximate the conditional density of the latent states given the observed data using sequential importance sampling (SIS) and resampling. SIS is the SMC method that forms the basis of the particle methods. However, sample degeneracy is typically a problem associated with the SIS method. Specifically, sample degeneracy occurs when all but one of the importance weights (as defined below) are close to zero. To avoid this problem, a resampling technique (e.g., bootstrapping) is applied to remove particles with small weights. The general concepts of particle filtering and smoothing for state-space models can be found in Kim & Stoffer (2008, pp. 828-829).

For the dynamic ZIP model, we implement the 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 $\mathbf{s}_{t|t-1}^{(i)} \sim \mathcal{N}_p(\boldsymbol{\Phi}\mathbf{s}_{t-1|t-1}^{(i)}, \boldsymbol{\Sigma})$ and $u_{t|t-1}^{(i)} \sim \text{Bernoulli}(\omega)$.

(F.2) Compute the filtering weights

$$q_{t|t-1}^{(i)} = \{(1 - u_{t|t-1}^{(i)})\lambda_{t|t-1}^{(i)}\}^{y_t} \exp\{-(1 - u_{t|t-1}^{(i)})\lambda_{t|t-1}^{(i)}\}/y_t!,$$

where $\log \lambda_{t|t-1}^{(i)} = \log w_t + \mathbf{x}_t^\top \boldsymbol{\beta} + z_{t|t-1}^{(i)}$ and $z_{t|t-1}^{(i)}$ is the first element of $\mathbf{s}_{t|t-1}^{(i)}$.

(F.3) Based on the above filtering weights, generate $(\mathbf{s}_{t|t}^{(i)}, u_{t|t}^{(i)})$ by resampling $(\mathbf{s}_{t|t-1}^{(i)}, u_{t|t-1}^{(i)})$ with replacement.

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

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

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

Next we use the Monte Carlo smoothing by Godsill et al. (2004) to obtain an approximate posterior sample of the latent variables given the observed data. In the particle smoothing step, we first choose $(\mathbf{s}_{n|n}^{(r)}, u_{n|n}^{(r)}) = (\mathbf{s}_{n|n}^{(i)}, u_{n|n}^{(i)})$ with probability $q_{n|n-1}^{(i)}$. Then for $t = n-1, \ldots, 1$:

(S.1) Calculate the smoothing weights

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

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

We obtain independent realizations by repeating the preceding processes for $r = 1, \ldots, R$. The forward-filtering and backward-smoothing procedure is the non-linear and non-Gaussian extension of the Kalman filtering and smoothing methods.

3.2.3 Monte Carlo EM Algorithm

To simplify the notation, we let $\mathbf{A}_t^{(j)}$, $\mathbf{b}_t^{(j)}$, $c_t^{(j)}$, $d_t^{(j)}$, and $e_t^{(j)}$ denote the conditional expectations of $\mathbf{s}_{t-1}\mathbf{s}_{t-1}^\top$, $z_t\mathbf{s}_{t-1}$, z_t^2, u_t, and $(1-u_t)\exp(z_t)$ evaluated at $\boldsymbol{\theta}^{(j)}$, respectively. In the Monte Carlo EM (MCEM) algorithm, we first compute

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

where the particle filtering and smoothing techniques are used to approximate the conditional expectations.

Given the state-space representation (3.3)-(3.5), the M-step of the MCEM algorithm is much easier than the direct maximization of the marginal likelihood of $y_{1:n}$. The following derivatives can be applied to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)})$:

$$\begin{aligned} \frac{\partial Q}{\partial\omega} &= \frac{1}{\omega}\sum_{t=1}^{n}d_t^{(j)} - \frac{1}{1-\omega}\sum_{t=1}^{n}(1-d_t^{(j)}), \\ \frac{\partial Q}{\partial\boldsymbol{\beta}} &= \sum_{t=1}^{n}\{(1-d_t^{(j)})y_t - e_t^{(j)}w_t\exp(\mathbf{x}_t^\top\boldsymbol{\beta})\}\mathbf{x}_t, \\ \frac{\partial Q}{\partial\boldsymbol{\phi}} &= \frac{1}{\sigma^2}\sum_{t=1}^{n}(b_t^{(j)} - \mathbf{A}_t^{(j)}\boldsymbol{\phi}), \\ \frac{\partial Q}{\partial\sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{t=1}^{n}(c_t^{(j)} - 2\boldsymbol{\phi}^\top\mathbf{b}_t^{(j)} + \boldsymbol{\phi}^\top\mathbf{A}_t^{(j)}\boldsymbol{\phi}). \end{aligned}$$

In the M-step, we update $\omega^{(j+1)}$, $\boldsymbol{\phi}^{(j+1)}$, and $\sigma^{(j+1)}$ as follows:

$$\begin{aligned}
\omega^{(j+1)} &= \frac{1}{n}\sum_{t=1}^{n} d_t^{(j)}, \\
\boldsymbol{\phi}^{(j+1)} &= \left(\sum_{t=1}^{n}\mathbf{A}_t^{(j)}\right)^{-1}\sum_{t=1}^{n}\mathbf{b}_t^{(j)}, \\
\sigma^{(j+1)} &= \sqrt{\frac{1}{n}\left\{\sum_{t=1}^{n} a_t^{(j)} - \left(\sum_{t=1}^{n}\mathbf{b}_t^{(j)}\right)^{\top}\left(\sum_{t=1}^{n}\mathbf{A}_t^{(j)}\right)^{-1}\sum_{t=1}^{n}\mathbf{b}_t^{(j)}\right\}}.
\end{aligned}$$

In addition, we can obtain $\boldsymbol{\beta}^{(j+1)}$ by fitting a weighted Poisson regression.

3.2.4 Observed Information Matrix

Once we obtain the MLE through the MCEM algorithm, we apply Louis' formula (Louis, 1982) to compute the observed information matrix $\mathbf{I}_o(\boldsymbol{\theta})$. Based on the missing information principle, we have

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

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

$$\begin{aligned}
\mathbf{I}_c(\boldsymbol{\theta}) &= \mathrm{E}\left(-\frac{\partial^2 l_c}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\middle| y_{1:n}\right), \\
\mathbf{I}_m(\boldsymbol{\theta}) &= \mathrm{E}\left(\frac{\partial l_c}{\partial\boldsymbol{\theta}}\frac{\partial l_c}{\partial\boldsymbol{\theta}^{\top}}\middle| y_{1:n}\right) - \mathrm{E}\left(\frac{\partial l_c}{\partial\boldsymbol{\theta}}\middle| y_{1:n}\right)\mathrm{E}\left(\frac{\partial l_c}{\partial\boldsymbol{\theta}^{\top}}\middle| y_{1:n}\right).
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial l_c}{\partial\omega} &= \frac{1}{\omega}\sum_{t=1}^{n} u_t - \frac{1}{1-\omega}\sum_{t=1}^{n}(1-u_t), \\
\frac{\partial l_c}{\partial\boldsymbol{\beta}} &= \sum_{t=1}^{n}(1-u_t)\{y_t - w_t\exp(z_t)\exp(\mathbf{x}_t^{\top}\boldsymbol{\beta})\}\mathbf{x}_t, \\
\frac{\partial l_c}{\partial\boldsymbol{\phi}} &= \frac{1}{\sigma^2}\sum_{t=1}^{n}(z_t - \boldsymbol{\phi}^{\top}\mathbf{s}_{t-1})\mathbf{s}_{t-1}, \\
\frac{\partial l_c}{\partial\sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{t=1}^{n}(z_t - \boldsymbol{\phi}^{\top}\mathbf{s}_{t-1})^2.
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial^2 l_c}{\partial\omega\partial\omega} &= -\frac{1}{\omega^2}\sum_{t=1}^{n} u_t - \frac{1}{(1-\omega)^2}\sum_{t=1}^{n}(1-u_t), \\
\frac{\partial^2 l_c}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -\sum_{t=1}^{n}(1-u_t)w_t \exp(z_t)\exp(\mathbf{x}_t^\top\boldsymbol{\beta})\mathbf{x}_t\mathbf{x}_t^\top, \\
\frac{\partial^2 l_c}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= -\frac{1}{\sigma^2}\sum_{t=1}^{n}\mathbf{s}_{t-1}\mathbf{s}_{t-1}^\top, \\
\frac{\partial^2 l_c}{\partial\sigma\partial\sigma} &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{t=1}^{n}(z_t - \boldsymbol{\phi}^\top\mathbf{s}_{t-1})^2, \\
\frac{\partial^2 l_c}{\partial\boldsymbol{\phi}\partial\sigma} &= -\frac{2}{\sigma^3}\sum_{t=1}^{n}(z_t - \boldsymbol{\phi}^\top\mathbf{s}_{t-1})\mathbf{s}_{t-1}.
\end{aligned}$$

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

3.3 Dynamic ZINB Model

In addition to zero-inflation, overdispersion could also be present for many count time series. In this section, we introduce a dynamic ZINB model for zero-inflated and overdispersed time series. Similar to the dynamic ZIP model, we write the dynamic ZINB model in the following state-space form:

$$\mathbf{s}_t|\mathbf{s}_{t-1} \sim \mathcal{N}_p(\boldsymbol{\Phi}\mathbf{s}_{t-1}, \boldsymbol{\Sigma}), \tag{3.7}$$

$$u_t \sim \text{Bernoulli}(\omega), \tag{3.8}$$

$$v_t \sim \text{Gamma}(k, 1/k), \tag{3.9}$$

$$y_t|\mathbf{s}_t, u_t, v_t \sim \text{Poisson}((1-u_t)v_t\lambda_t), \tag{3.10}$$

where k is the dispersion parameter of the NB distribution. Here, we let $\boldsymbol{\theta} = (\omega, k, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \sigma)^\top$ denote the vector of unknown parameters. Figure 3.3 illustrates the dynamic ZINB model that is defined by (3.7)-(3.10).

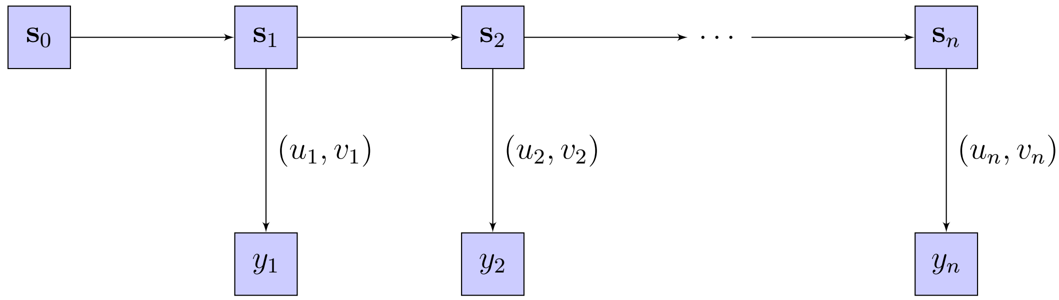

Figure 3.3: Graphical illustration of the state evolution and data generation in the dynamic ZINB model.

The dynamic ZINB model is very general and it includes many important models as special cases (see Table 3.1). Specifically, it reduces to

- dynamic NB regression when $\omega = 0$,
- dynamic ZIP regression when $1/k \to 0$,
- dynamic Poisson regression when $\omega = 0$ and $1/k \to 0$.

In the case where $z_t = 0$, the dynamic Poisson, NB, ZIP, and ZINB regressions simply reduce to the four ordinary regression models for independent data.

Table 3.1: Parameter-driven models for zero-inflated and overdispersed time series.

Zero-Inflation	Overdispersion	Autocorrelation	Model
No	No	No	Poisson Regression
No	No	Yes	Dynamic Poisson Regression
No	Yes	No	NB Regression
No	Yes	Yes	Dynamic NB Regression
Yes	No	No	ZIP Regression
Yes	No	Yes	Dynamic ZIP Regression
Yes	Yes	No	ZINB Regression
Yes	Yes	Yes	Dynamic ZINB Regression

Based on the state-space representation, we can orthogonally decompose the complete-data likelihood (i.e., the joint density of $\mathbf{s}_{0:n}$, $u_{1:n}$, $v_{1:n}$, and $y_{1:n}$) as follows:

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

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

$$\begin{aligned}
l_c(\boldsymbol{\theta}) &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(z_t - \boldsymbol{\phi}^\top \mathbf{s}_{t-1})^2 \\
&\quad + \sum_{t=1}^{n}\{u_t \log\omega + (1-u_t)\log(1-\omega)\} \\
&\quad + \sum_{t=1}^{n}\{k\log k - \log\Gamma(k) + k(\log v_t - v_t)\} \\
&\quad + \sum_{t=1}^{n}(1-u_t)\{y_t \mathbf{x}_t^\top \boldsymbol{\beta} - v_t w_t \exp(z_t)\exp(\mathbf{x}_t^\top \boldsymbol{\beta})\}.
\end{aligned}$$

In the MCEM algorithm, we first compute

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) &= \mathrm{E}\{l_c(\boldsymbol{\theta})|y_{1:n}, \boldsymbol{\theta}^{(j)}\} \\
&= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(c_t - 2\boldsymbol{\phi}^\top \mathbf{b}_t + \boldsymbol{\phi}^\top \mathbf{A}_t \boldsymbol{\phi}) \\
&\quad + \sum_{t=1}^{n}\{d_t \log\omega + (1-d_t)\log(1-\omega)\} \\
&\quad + \sum_{t=1}^{n}\{k\log k - \log\Gamma(k) + k(f_t - e_t)\} \\
&\quad + \sum_{t=1}^{n}\{(1-d_t) y_t \mathbf{x}_t^\top \boldsymbol{\beta} - g_t w_t \exp(\mathbf{x}_t^\top \boldsymbol{\beta})\},
\end{aligned}$$

where $\mathbf{A}_t$, $\mathbf{b}_t$, c_t, d_t, e_t, f_t and g_t are the conditional expectations of $\mathbf{s}_{t-1}\mathbf{s}_{t-1}^\top$, $z_t\mathbf{s}_{t-1}$, z_t^2, u_t, v_t, $\log v_t$ and $(1-u_t)v_t\exp(z_t)$, respectively. The following derivatives can be applied to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)})$:

$$
\begin{aligned}
\frac{\partial Q}{\partial \omega} &= \frac{1}{\omega}\sum_{t=1}^{n} d_t - \frac{1}{1-\omega}\sum_{t=1}^{n}(1-d_t), \\
\frac{\partial Q}{\partial k} &= n\{1+\log k - \Psi_0(k)\} + \sum_{t=1}^{n}(f_t - e_t), \\
\frac{\partial Q}{\partial \boldsymbol{\beta}} &= \sum_{t=1}^{n}\{(1-d_t)y_t - g_t w_t \exp(\mathbf{x}_t^\top \boldsymbol{\beta})\}\mathbf{x}_t, \\
\frac{\partial Q}{\partial \boldsymbol{\phi}} &= \frac{1}{\sigma^2}\sum_{t=1}^{n}(\mathbf{b}_t - \mathbf{A}_t\boldsymbol{\phi}), \\
\frac{\partial Q}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{t=1}^{n}(c_t - 2\boldsymbol{\phi}^\top \mathbf{b}_t + \boldsymbol{\phi}^\top \mathbf{A}_t \boldsymbol{\phi}),
\end{aligned}
$$

where $\Psi_0(\cdot)$ is the digamma function. Similar to the dynamic ZIP model, we have closed-form solutions for $\omega^{(j+1)}$, $\boldsymbol{\phi}^{(j+1)}$, and $\sigma^{(j+1)}$. In addition, we obtain $k^{(j+1)}$ and $\boldsymbol{\beta}^{(j+1)}$ by iterative algorithms.

In general, identifiability problems could arise as a model becomes increasingly complicated. For the dynamic ZINB model, identifiability could be an issue since z_t and v_t can both account for overdispersion in the data. To investigate the identifiability issue, we consider a simple dynamic ZINB model with $\omega = 0.2$, $1/k = 0.5$, $\beta = 2$, $\boldsymbol{\phi} = \mathbf{0}$, and $\sigma = 0.5$. Note that the generating model may be viewed as a ZINB mixed model since there is no autocorrelation in the random effects z_t. We first simulate a dataset of length 100 from the ZINB mixed model and then attempt to estimate the parameters $1/k$ and σ from the data. Figure 3.4 displays the likelihood surface of $1/k$ and σ in the ZINB mixed model. It is clear that the likelihood function is log-concave, indicating the parameters $1/k$ and σ are actually identifiable. However, the estimated parameters (identified by the red dot) are quite different from the true parameters (identified by the green dot). Thus, we do not recommend fitting a dynamic ZINB model when sample information is limited or only a short sequence of observations is available.

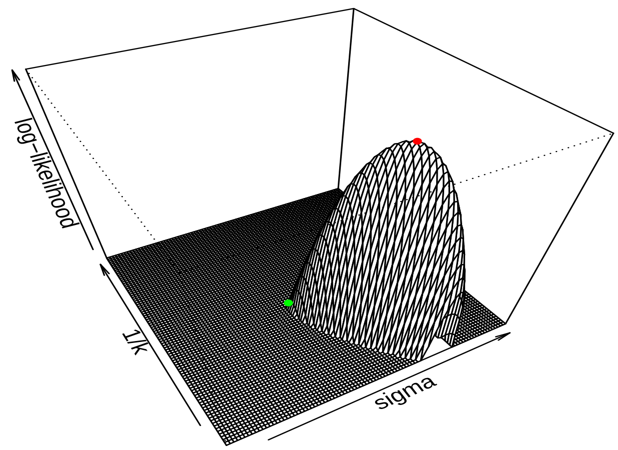

Figure 3.4: Likelihood surface for the ZINB mixed model.

3.4 Simulated Examples

In this section, we consider time series data that are simulated from four different dynamic models (Poisson, NB, ZIP, and ZINB). All these models belong to the class of general dynamic ZINB models. For each of the four models, we assume $\{z_t\}$ is an AR(2) process such that

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \epsilon_t,$$

where ϵ_t is a white noise process with mean 0 and standard deviation $\sigma = 0.5$. We choose $\phi_1 = 0.8$ and $\phi_2 = -0.6$ to ensure that the AR(2) process is stationary (Cryer & Chan, 2008, pp. 71-72). Conditioning on the current state z_t, we assume the observation y_t has a ZINB distribution with p.m.f. given by

$$f_{Y_t}(y_t|z_t;\omega,k,\lambda_t)=\begin{cases}\omega+(1-\omega)\left(\dfrac{k}{k+\lambda_t}\right)^k, & \text{if } y_t=0,\\[2ex] (1-\omega)\dfrac{\Gamma(k+y_t)}{\Gamma(k)y_t!}\left(\dfrac{k}{k+\lambda_t}\right)^k\left(\dfrac{\lambda_t}{k+\lambda_t}\right)^{y_t}, & \text{if } y_t>0,\end{cases}$$

where we have $\log\lambda_t=\beta+z_t$ and β is set to be 2 for all the models. The following lists the true parameters in the four generating models.

- Dynamic Poisson Model

  $\omega=0$, $1/k=0$, $\beta=2$, $\phi_1=0.8$, $\phi_2=-0.6$, and $\sigma=0.5$.

- Dynamic NB Model

  $\omega=0$, $1/k=0.5$, $\beta=2$, $\phi_1=0.8$, $\phi_2=-0.6$, and $\sigma=0.5$.

- Dynamic ZIP Model

  $\omega=0.2$, $1/k=0$, $\beta=2$, $\phi_1=0.8$, $\phi_2=-0.6$, and $\sigma=0.5$.

- Dynamic ZINB Model

  $\omega=0.2$, $1/k=0.5$, $\beta=2$, $\phi_1=0.8$, $\phi_2=-0.6$, and $\sigma=0.5$.

The sample size (i.e., length of the series) is set to be 200. Table 3.2 presents the dispersion indices (i.e., variance-to-mean ratios) and proportions of zeros for the simulated time series.

Table 3.2: Dispersion indices and proportions of zeros for the simulated examples.

	Poisson	NB	ZIP	ZINB
Dispersion Index	5.202	11.394	8.905	17.603
Proportion of Zeros	0.005	0.075	0.255	0.300

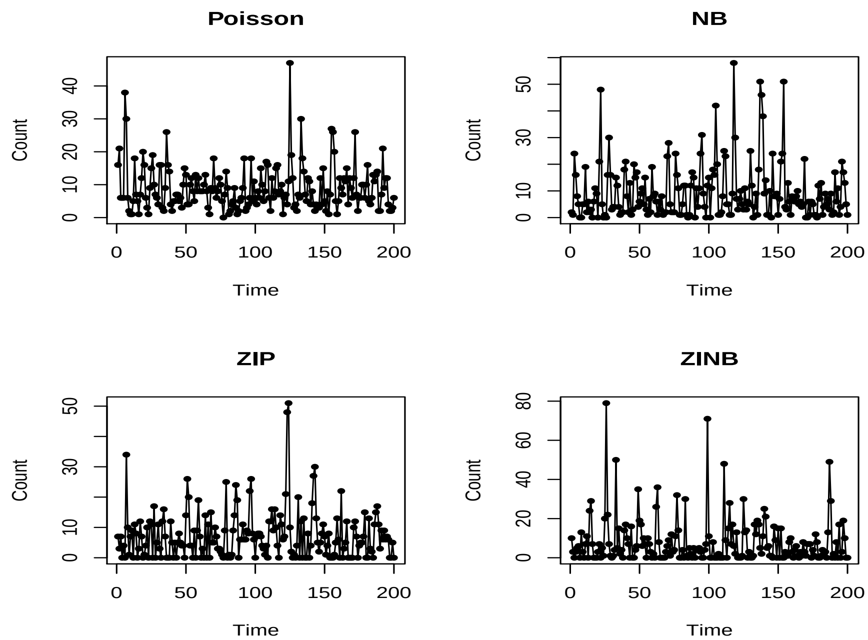

Figure 3.5: Time series plots for the simulated examples.

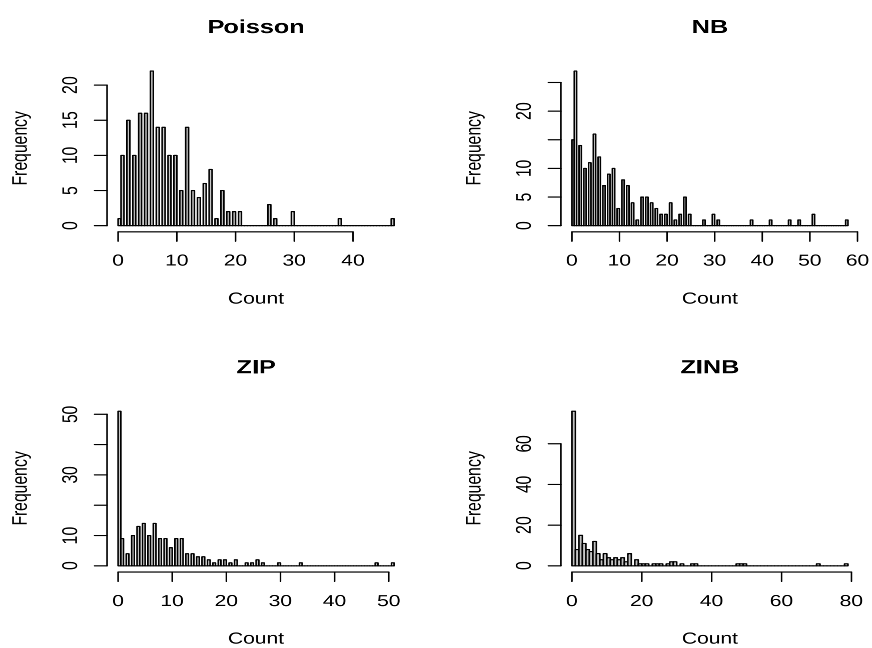

Figure 3.6: Histograms for the simulated examples.

Figure 3.5 displays the time series plots for the four time series of counts. All four series generated from the dynamic models are stationary over time. The corresponding histograms of the simulated data are presented in Figure 3.6. A bimodal shape of the density is observed for data generated from the dynamic ZIP and ZINB models.

Based on Figure 3.7, the partial autocorrelation functions (ACFs) would suggest an AR(1) structure for data generated from dynamic ZIP and ZINB models. Since the true order of the AR structure is 2, the traditional model specification tools (e.g., ACF and partial ACF) fail to work here and can be very misleading in the presence of zero-inflation.

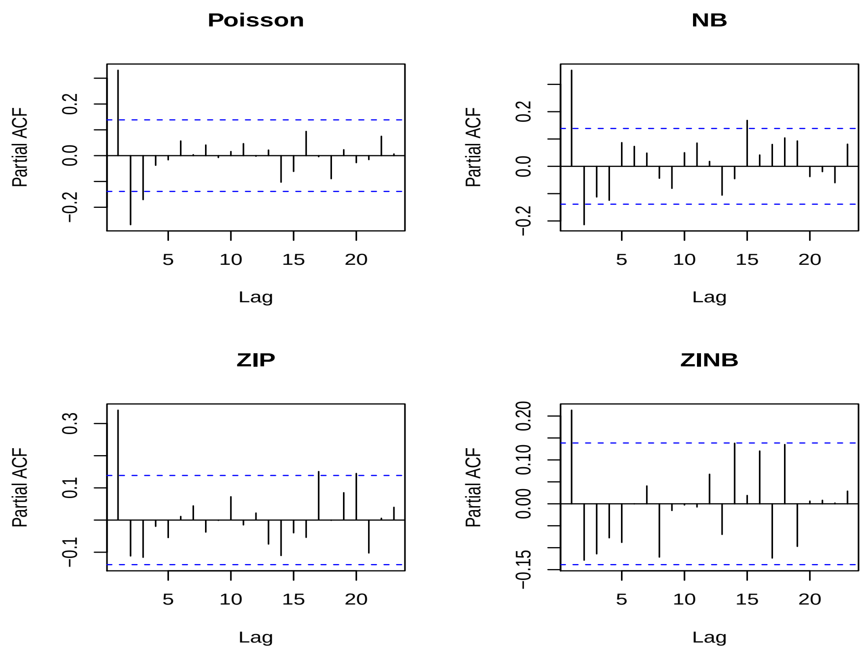

Figure 3.7: Partial ACF plots for the simulated examples.

Table 3.3 presents the true and estimated parameters of the generating models. In general, the absolute differences between the true and estimated parameters are small. Smaller differences could be obtained by increasing the sample size.

Table 3.3: True and estimated parameters for the simulated examples.

	ω	$1/k$	β	ϕ_1	ϕ_2	σ
True	0.200	0.500	2.000	0.800	-0.600	0.500
Poisson			2.065	0.772	-0.557	0.441
NB		0.476	2.064	0.833	-0.663	0.492
ZIP	0.243		2.032	0.882	-0.565	0.466
ZINB	0.257	0.425	2.098	0.622	-0.402	0.659

Figure 3.8 display the trace plots of the log-likelihood for the four dynamic models. We can see that the log-likelihood increases dramatically at the very beginning, and then stabilizes and fluctuates after a certain number of iterations.

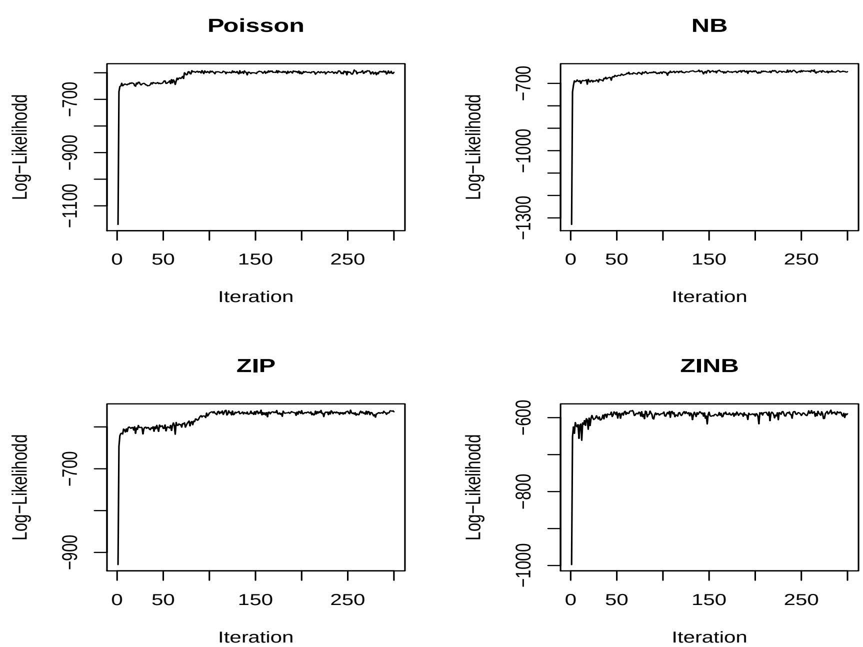

Figure 3.8: Trace plots of the log-likelihood for the simulated examples.

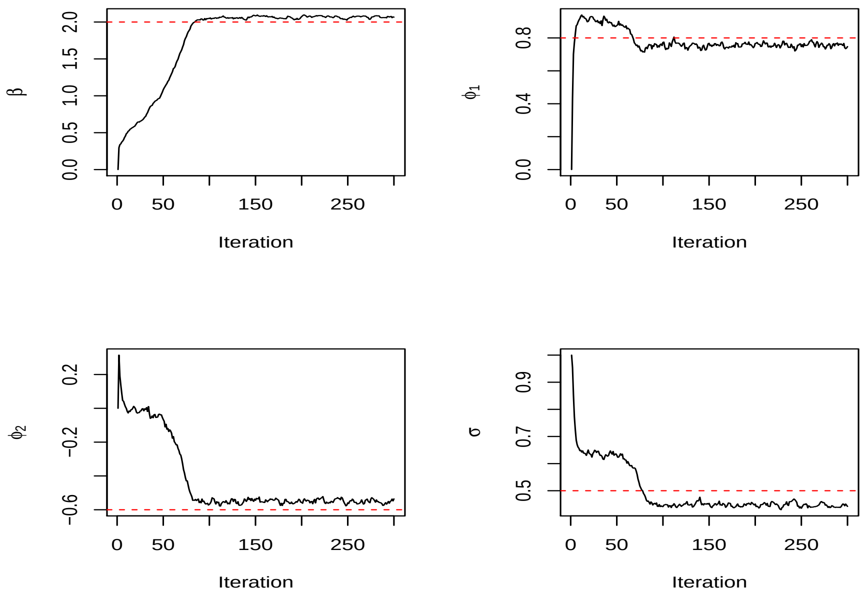

Figure 3.9: Trace plots of the estimated parameters for the Poisson time series.

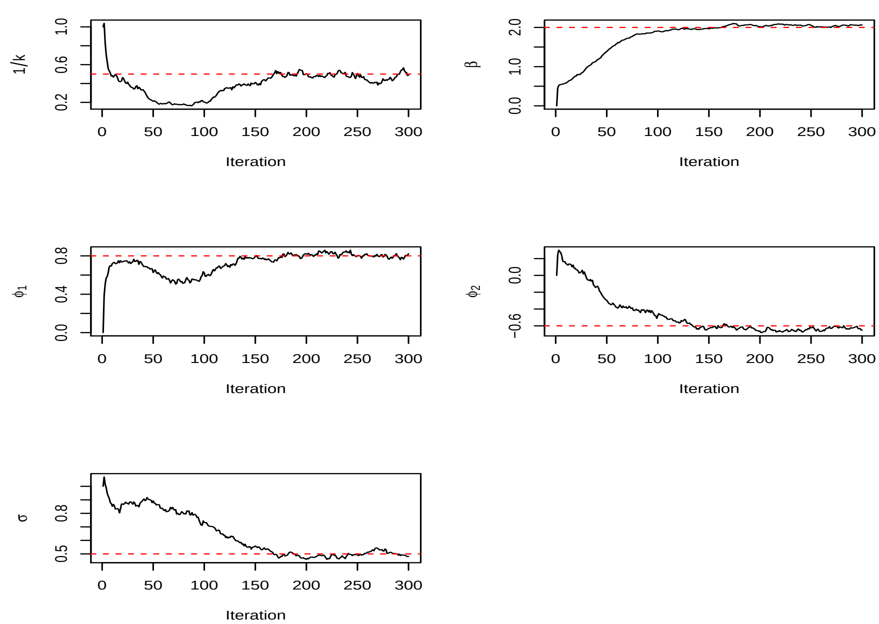

Figure 3.10: Trace plots of the estimated parameters for the NB time series.

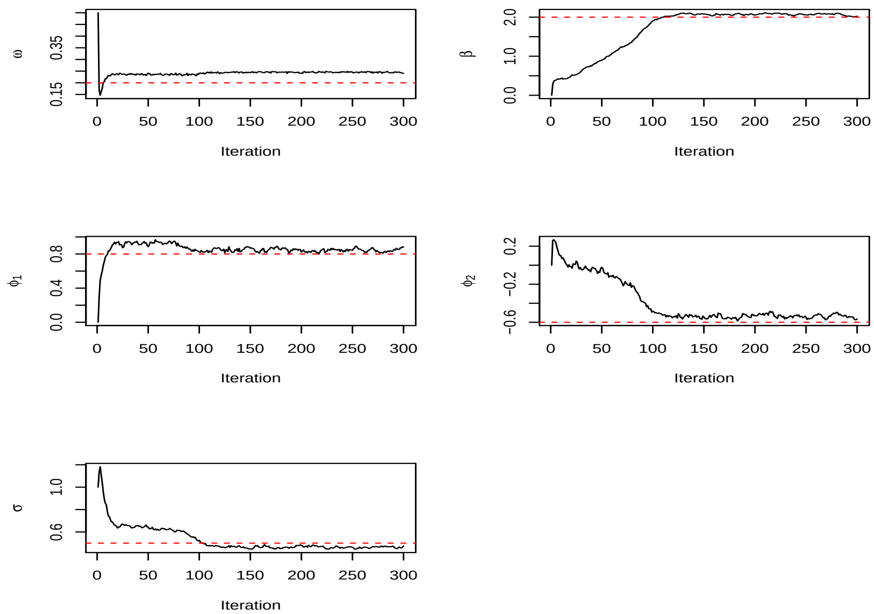

Figure 3.11: Trace plots of the estimated parameters for the ZIP time series.

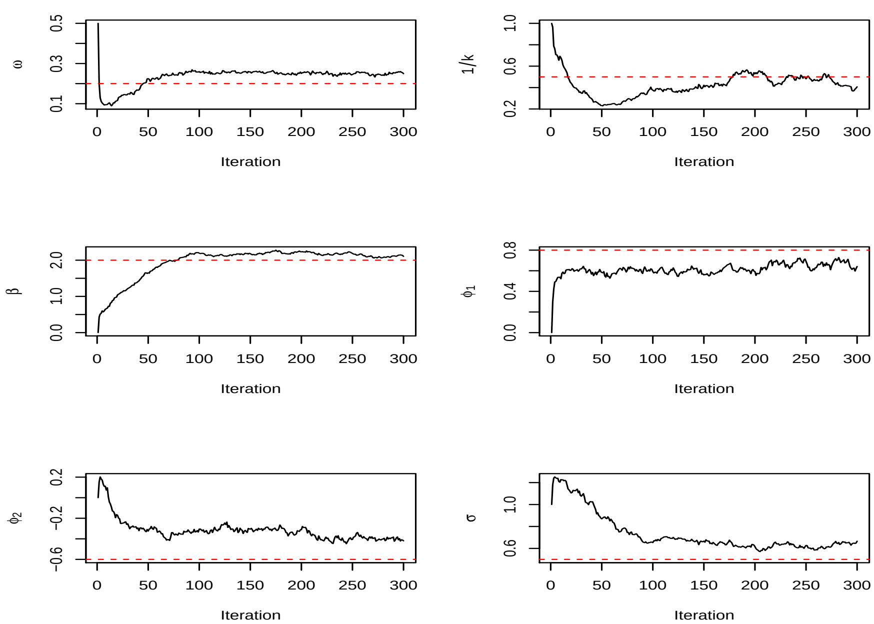

Figure 3.12: Trace plots of the estimated parameters for the ZINB time series.

The trace plots of the four dynamic models are presented in Figures 3.9-3.12. We notice that some of the parameters are still changing substantially even when the log-likelihood is stabilized. For the MCEM algorithm, a stopping rule solely based on the log-likelihood change is unreliable. Thus, we strongly recommend checking the trace plots of the parameters to ensure the model is truly stationary. For the dynamic Poisson, NB, and ZIP models, all of the estimated parameters approach the true parameters within a reasonable tolerance level. However, the estimated parameters for ϕ_1, ϕ_2, and σ in the dynamic ZINB model are quite different from the true values (Figure 3.12). This phenomenon agrees the identifiability issue that has been discussed at the end of Section 3.3.

3.5 Application

We revisit the application pertaining to public health surveillance for syphilis in this section. We fit four dynamic models (Poisson, NB, ZIP, and ZINB) to the syphilis data from Maryland for 2007 to 2010. An AR(1) correlation structure is employed in all models; higher order AR structures do not help improve the model fit. The Monte Carlo EM algorithm is used to fit the models.

Table 3.4: Dynamic ZIP and Poisson models for the syphilis data from Maryland.

	ZIP (AIC = 932.9)			Poisson (AIC = 1009.1)		
	Estimate	SE	P-Value	Estimate	SE	P-Value
β_0	1.6882	0.0763	<0.0001	1.4031	0.1306	<0.0001
β_1	-1.6949	0.7050	0.0162	-4.3199	1.1655	0.0002
ϕ_1	-0.1174	0.1326	0.3760	0.2051	0.0940	0.0292
σ	0.2540			0.7710		
ω	0.2689					

Table 3.5: Dynamic ZINB and NB models for the syphilis data from Maryland.

	ZINB (AIC = 934.6)			NB (AIC = 993.5)		
	Estimate	SE	P-Value	Estimate	SE	P-Value
β_0	1.6886	0.0764	<0.0001	1.6318	0.1201	<0.0001
β_1	-1.4571	0.6966	0.0365	-3.8968	1.0394	0.0002
ϕ_1	-0.0431	0.1321	0.7440	0.0705	0.1343	0.5998
σ	0.1473			0.1368		
$1/k$	0.0332			0.7030		
ω	0.2708					

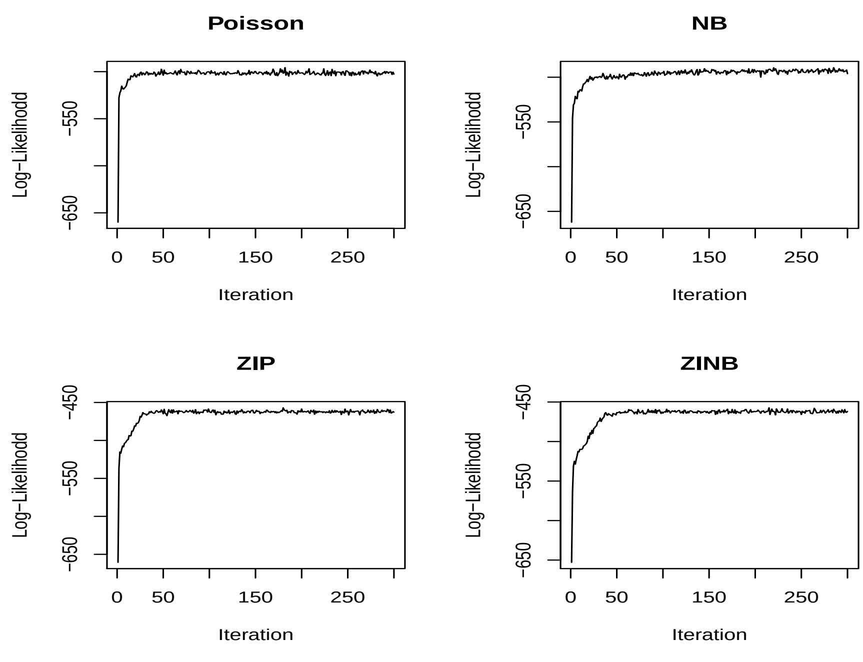

Figure 3.13: Trace plots of the log-likelihood for models fit to the syphilis data from Maryland.

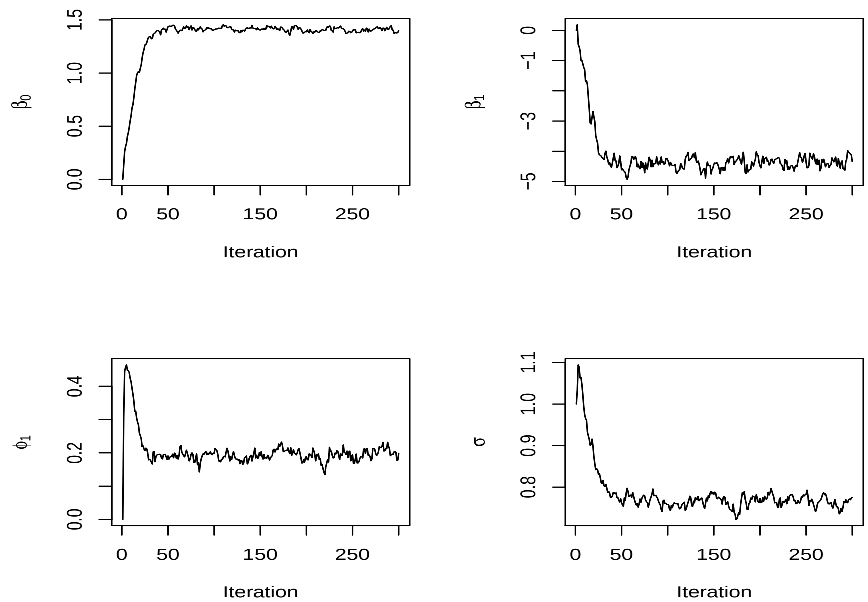

Figure 3.14: Trace plots of the estimated parameters for the dynamic Poisson model.

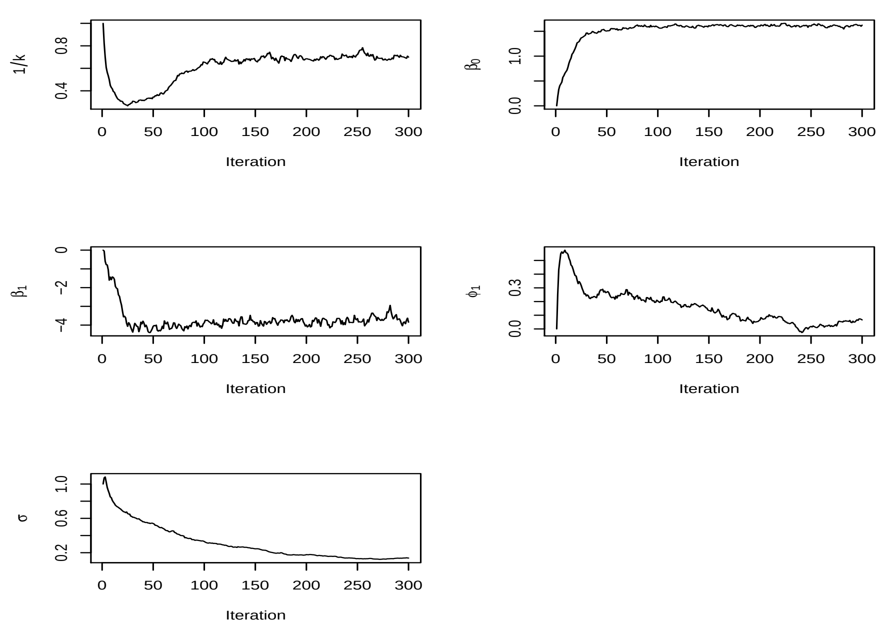

Figure 3.15: Trace plots of the estimated parameters for the dynamic NB model.

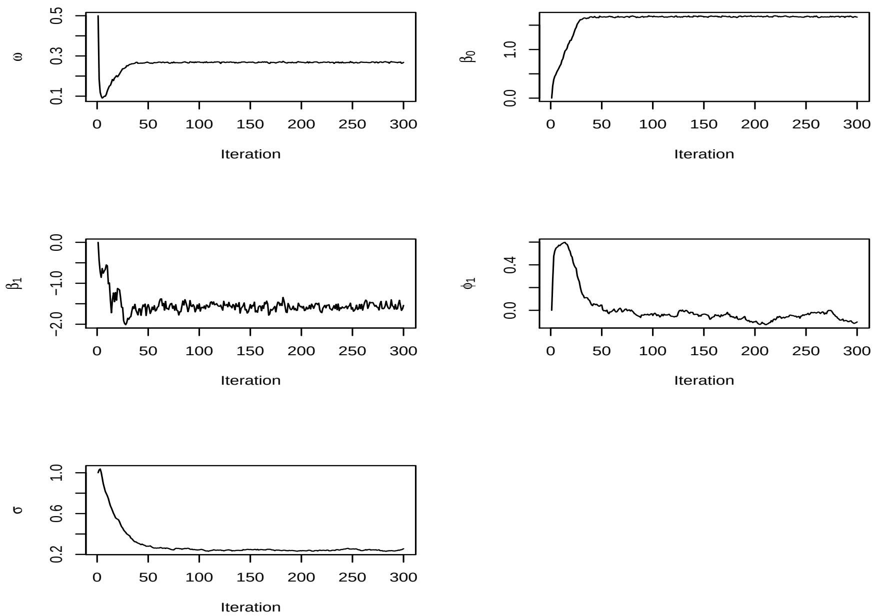

Figure 3.16: Trace plots of the estimated parameters for the dynamic ZIP model.

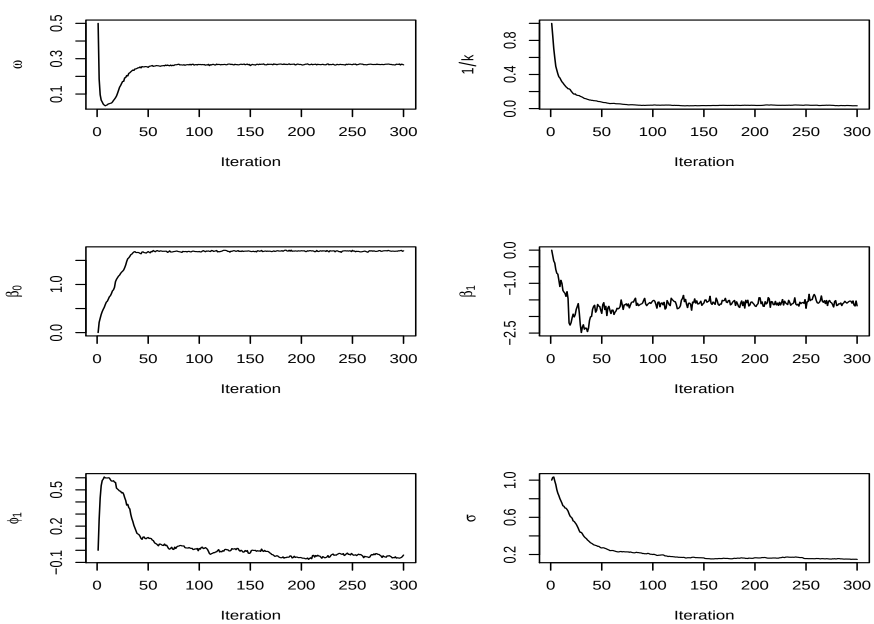

Figure 3.17: Trace plots of the estimated parameters for the dynamic ZINB model.

Unlike the exact EM algorithm, the log-likelihood in the MCEM algorithm is not guaranteed to increase at each iteration. Instead, the log-likelihood will increase dramatically in the first several iterations, and then stabilize as the estimated parameters become very close to the MLE (Figure 3.13). The trace plots of the four dynamic models are presented in Figures 3.14-3.17. Table 3.4 displays the regression output for the dynamic ZIP model and its Poisson counterpart. The regression output for the dynamic ZINB model and its NB counterpart is presented in Table 3.5. The p-values are not reported for σ, ω, and $1/k$ in Tables 3.4 and 3.5, as traditional asymptotic theories fail to work when the hypothesis testing is conducted on the boundary of the parameter space. Based on the AIC, the dynamic ZIP and ZINB models significantly outperform their Poisson and NB counterparts, which fail to account for zero-inflation in the data.

CHAPTER 4
THE R PACKAGE ZIM

We have developed an R package, called ZIM (Zero-Inflated Models), to analyze count time series with excess zeros. The purpose of this chapter is to provide a general introduction to the ZIM package. Throughout the chapter, the syphilis data from Maryland will be repeatedly used as an example. The structure of this chapter is as follows. Section 4.1 first introduces the R functions for time series models based on the ZIP and ZINB distributions. The usage of those functions to fit observation-driven models and parameter-driven models is then discussed in Section 4.2 and Section 4.3, respectively.

4.1 Overview of the Package

In the ZIM package, the following are the main functions to implement the ZIP and ZINB autoregressions that have been proposed in Chapter 2.

- `zim`
- `zim.fit`
- `zim.control`

The function `zim` is a user-friendly function to fit zero-inflated observation-driven models. Its usage is very similar to that of the well-known function `glm`. Here, `zim.fit` is the fitter function and it is called by `zim` to fit the models. The `zim.fit` function should not be used directly unless by experienced users. The function `zim.control` is an auxiliary function for `zim` fitting. It is typically only used internally by `zim.fit`, but may be used to construct a `control` argument to either function.

Compared to observation-driven models, parameter estimation in parameter-driven models is much more challenging. The following are the functions that can be used to fit the dynamic ZIP and ZINB models that have been proposed in Chapter 3. In addition, the dynamic Poisson and NB models can also be fit since they are simply special cases of the dynamic zero-inflated models.

- `dzim`
- `dzim.fit`
- `dzim.filter`
- `dzim.smooth`
- `dzim.control`

The function `dzim` is a user-friendly function to fit dynamic zero-inflated models. The default order for the autoregressive process is assumed to be one. Here, `dzim.fit` is the fitter function and it is called by `dzim` to fit the models. The function `dzim.control` is an auxiliary function for `dzim` fitting. The functions `dzim.filter` and `dzim.smooth` are used to implement the particle filtering and particle smoothing methods, respectively.

4.2 Observation-Driven Models

4.2.1 ZIP Autoregression

We first fit a ZIP autoregression with an AR(1) correlation structure. The linear trend is included in both the log-linear and logistic parts of the model.

```
> library(ZIM)
> data(syph)
> count <- syph$a33
> ar1 <- bs(count > 0)
> trend <- 1:length(count) / 1000
> zim(count ~ ar1 + trend | trend)
```

```
Call:
zim(formula = count ~ ar1 + trend | trend)

Coefficients (log-linear):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.48942    0.11995 12.4175  < 2e-16 ***
ar1          0.22111    0.10072  2.1954  0.02813 *
trend       -1.01004    0.66687 -1.5146  0.12987

Coefficients (logistic):
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) -1.93321    0.37196 -5.1974 2.021e-07 ***
trend        8.60517    2.80827  3.0642  0.002182 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Test for overdispersion (H0: ZIP vs. H1: ZINB)
score.test: 2.6031
p.value: 0.0046196

Criteria for assessing goodness of fit
loglik: -454.3903
aic: 918.7806
bic: 935.4683
tic: 920.7761

Number of EM-NR iterations: 11
```

```
Maximum absolute gradient: 1.021405e-14
```

The EM-NR algorithm is used as the default algorithm in the `zim` function. The score test for overdispersion suggests that the ZINB model could provide a better fit to the syphilis data (p = 0.0046).

4.2.2 ZINB Autoregression

As suggested by the score test, we next fit a ZINB autoregression, with all the other components remaining the same as in the ZIP autoregression.

```
> zim(count ~ ar1 + trend | trend, dist = "zinb")

Call:
zim(formula = count ~ ar1 + trend | trend, dist = "zinb")


Coefficients (log-linear):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.47240    0.13873 10.6132  < 2e-16 ***
ar1          0.23164    0.11522  2.0105  0.04438 *
trend       -1.00364    0.77154 -1.3008  0.19332


Coefficients (logistic):
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) -1.97940    0.38563 -5.1329 2.853e-07 ***
trend        8.71684    2.88697  3.0194  0.002533 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for negative binomial taken to be 15.4711)
```

```
Criteria for assessing goodness of fit
loglik: -451.7464
aic: 915.4927
bic: 935.5179
tic: 915.974

Number of EM-NR iterations: 11
Maximum absolute gradient: 5.087796e-08
```

The AIC and TIC suggest a marginal improvement when the ZINB autoregression is used. However, the BIC values for the ZIP and ZINB autoregressions are not distinguishable. This should not be surprising as BIC tends to penalize more for complexity.

4.3 Parameter-Driven Models

4.3.1 Dynamic ZIP Model

We now fit a dynamic ZIP model to the syphilis data. The trend is included as a deterministic covariate in the log-linear model. The zero-inflation parameter is assumed to be constant over time.

```
> dzim(count ~ trend, dist = "zip", minit = 300)

Call:
dzim(formula = count ~ trend, dist = "zip", minit = 300)

(Zero-inflation parameter taken to be 0.2689)

Coefficients (log-linear):
```

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.688213   0.076337 22.1152  < 2e-16 ***
trend       -1.694876   0.704988 -2.4041  0.01621 *

Coefficients (autoregressive):
    Estimate Std. Error z value Pr(>|z|)
ar1 -0.11739    0.13260 -0.8853    0.376
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Standard deviation parameter taken to be 0.254)

Criteria for assessing goodness of fit
loglik: -461.425
aic: 932.8501
bic: 949.5617
tic: 947.0466
```

4.3.2 Dynamic ZINB Model

We next fit a dynamic ZINB model to see whether a need remains for the NB dispersion parameter.

```
> dzim(count ~ trend, dist = "zinb", minit = 300)

Call:
dzim(formula = count ~ trend, dist = "zinb", minit = 300)

(Zero-inflation parameter taken to be 0.2708)
```

```
(Dispersion parameter for negative binomial taken to be 30.1071)

Coefficients (log-linear):
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.688631   0.076437 22.0919  < 2e-16 ***
trend       -1.457090   0.696551 -2.0919  0.03645 *

Coefficients (autoregressive):
     Estimate Std. Error z value Pr(>|z|)
ar1 -0.043149   0.132108 -0.3266    0.744
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Standard deviation parameter taken to be 0.1473)

Criteria for assessing goodness of fit
loglik: -461.2986
aic: 934.5971
bic: 954.6511
tic: 956.493
```

Since the inclusion of the correlated random effect z_t can account for both autocorrelation and overdispersion, fitting a more complicated ZINB model does not help to further reduce AIC, BIC, or TIC.

CHAPTER 5
CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

Count time series with excess zeros are often encountered in many biomedical and public health applications. Although the Poisson and negative binomial distributions have been widely used in practice for discrete count data, their forms are too simplistic to accommodate zero-inflation. Failure to account for zero-inflation while analyzing such data may result in misleading inferences and the detection of spurious associations.

Regression models based on the ZIP and ZINB distributions have been well established for data that are independently distributed. These zero-inflated regression models have been extended to analyze longitudinal and multilevel data. However, statistical models for zero-inflated time series are lacking in the literature. In this thesis we propose two classes of statistical models to analyze time series of counts containing extra zeros. An R package has been developed to fit both observation-driven and parameter-driven models. Both simulated and real examples are presented to illustrate the proposed methodologies.

5.2 Future Directions

There are a number of extensions that can enhance our work in this thesis. We outline several important future directions.

- We have developed models for zero-inflated time series and have applied them to the temporal surveillance of syphilis. From a public health perspective, it would also be of crucial importance to have models that can be used for syphilis surveillance at different geographical locations. To accomplish this, zero-inflated spatial or spatio-temporal models need to be developed.

- In both observation-driven and parameter-driven models, we only consider autoregressive terms to model the temporal correlation. In some cases, it may be desirable to also include moving average components to account for more complicated correlation structures in the data.

- The focus of this thesis is on univariate count time series with excess zeros. In practice, the data could be zero-inflated and also multivariate in nature. Thus, statistical models based on multivariate zero-inflated distributions could be developed to accommodate simultaneous zero-inflation in such settings.

- In this thesis, we have applied different model selection criteria to determine the autoregressive order in the count series. It would also be helpful to develop some diagnostic tools to identify the correlation structure, analogous to the ACF and PACF for traditional ARMA processes.

APPENDIX

In this appendix we provide a proof for Theorem 2 presented in Chapter 2. At any time t, we let

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

and

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

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

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

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

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

and

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

Moreover, it can be easily verified that

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

and

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

Thus, we have

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

and

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

Combining the preceding equations 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

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

Applying the delta method to Theorem 1, we have

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

and

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

This completes the proof of Theorem 2.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Bosch, R. J. and Ryan, L. M. (1998). Generalized Poisson Models Arising from Markov Processes. *Statistics & Probability Letters*, **39**, 205-212.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting, and Control*. Oakland, CA: Holden-Day.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, **87**, 493-500.

Cavanaugh, J. E. and Shumway, R. H. (1996). On computing the expected Fisher information matrix for state-space model parameters. *Statistics and Probability Letters*, **26**, 347-355.

Cavanaugh, J. E. (1997). Unifying the derivations of the Akaike and corrected Akaike information criteria. *Statistics and Probability Letters*, **31**, 201-208.

Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, **90**, 242-252.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62**, 269-276.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, **8**, 93-115.

Cryer, J. D. and Chan, K. S. (2008). *Time Series Analysis: With Applications in R, Second Edition*. New York, NY: Springer.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika*, **90**, 777-790.

Davis, R. A. and Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika*, **96**, 735-749.

Dean, C. and Lawless, J. F. (1989). Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association*, **84**, 467-472.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1-39.

Diggle, P. J., Heagerty, P., Liang K.-Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data, Second Edition*. Oxford: Oxford University Press.

Durbin, J. and Koopman, S. J. (2000). 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**, 3-56.

Fokianos, K. and Kedem, B. (1998). Prediction and classification of non-stationary categorical time series. *Journal of Multivariate Analysis*, **67**, 277-296.

Fokianos, K. and Kedem, B. (2004). Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis*, **25**, 173-197.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, **104**, 1430-1439.

Fokianos, K. and Tjøstheim, D. (2011). Log-linear Poisson autoregression. *Journal of Multivariate Analysis*, **102**, 563-578.

Garay, A. M., Hashimoto, E. M., Ortega, E. M. M., and Lachos, V. H. (2011). On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics and Data Analysis*, **55**, 1304-1318.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, **56**, 1030-1039.

Hall, D. B. and Berenhaut, K. S. (2002). Score tests for heterogeneity and overdispersion in zero-inflated Poisson and binomial regression models. *The Canadian Journal of Statistics*, **30**, 415-430.

Hall, D. B. and Zhang, Z. (2004). Marginal models for zero inflated clustered data. *Statistical Modelling*, **4**, 161-180.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. New York: Academic Press.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and The Kalman Filter*. Cambridge University Press.

Kaufmann, H. (1987). Regression models for nonstationary categorical time series: Asymptotic estimation theory. *The Annals of Statistics*, **15**, 79-98.

Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*. New Jersey: Wiley.

Kim, J. and Stoffer, D. S. (2008). Fitting stochastic volatility models in the presence of irregular sampling via particle methods and the EM algorithm. *Journal of Time Series Analysis*, **29**, 811-833.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1-14.

Lee, A. H., Wang, K., Scott, J. A., Yau, K. K. W., and McLachlan, G. J. (2006). Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medical Research*, **15**, 47-61.

Lee, A. H., Wang, K., and Yau, K. K. W. (2001). Analysis of zero-inflated Poisson data incorporating extent of exposure. *Biometrical Journal*, **43**, 963-975.

Liu, H. and Chan, K. S. (2010). Introducing COZIGAM: An R Package for Unconstrained and Constrained Zero-Inflated Generalized Additive Model Analysis. *Journal of Statistical Software*, **35**, 1-26.

Loeys, T., Moerkerke, B., De Smet, O. and Buysse, A. (2011). The analysis of zero-inflated count data: Beyond zero-inflated Poisson regression. *British Journal of Mathematical and Statistical Psychology*, **43**, 963-975.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, **33**, 341-365.

Nelson, K. P. and Leroux, B. G. (2006). Statistical models for autocorrelated count data. *Statistics in Medicine*, **25**, 1413-1430.

Oh, M. S. and Lim, Y. B. (2001). Bayesian analysis of time series Poisson data. *Journal of Applied Statistics*, **28**, 259-271.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, **57**, 120-125.

Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605-610.

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, **3**, 253-264.

Shumway, R. H. and Stoffer, D. S. (2006). *Time Series Analysis and Its Application: With R Examples (2nd ed.)*. New York: Springer.

Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Science*, **153**, 12-18.

Van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics*, **51**, 738-743.

Wang, P. (2001). Markov zero-inflated Poisson regression models for a time series of counts with excess zeros. *Journal of Applied Statistics*, **28**, 623-632.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, **80**, 73-83.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association*, **76**, 419-433.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1-25.

Yau, K. K. W., Lee, A. H., and Carrivick, P. J. W. (2004). Modeling zero-inflated count series with application to occupational health. *Computer Methods and Programs in Biomedicine*, **74**, 47-52.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, **75**, 621-629.

Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, **44**, 1019-1031.

Zeileis, A., Kleiber, C., and Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, **27**, 1-25.

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