Statistical Modelling 2015; **15**(**1**): 70–90

State-space models for count time series with excess zeros

Ming Yang[1], **Joseph E Cavanaugh**[2] **and Gideon KD Zamba**[2]
[1] Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA
[2] Department of Biostatistics, College of Public Health, University of Iowa, USA

Abstract: Count time series are frequently encountered in biomedical, epidemiological and public health applications. In principle, such series may exhibit three distinctive features: overdispersion, zero-inflation and temporal correlation. Developing a modelling framework that is sufficiently general to accommodate all three of these characteristics poses a challenge. To address this challenge, we propose a flexible class of dynamic models in the state-space framework. Certain models that have been previously introduced in the literature may be viewed as special cases of this model class. For parameter estimation, we devise a Monte Carlo Expectation-Maximization (MCEM) algorithm, where particle filtering and particle smoothing methods are employed to approximate the high-dimensional integrals in the E-step of the algorithm. To illustrate the proposed methodology, we consider an application based on the evaluation of a participatory ergonomics intervention, which is designed to reduce the incidence of workplace injuries among a group of hospital cleaners. The data consists of aggregated monthly counts of work-related injuries that were reported before and after the intervention.

Key words: autocorrelation; interrupted time series; intervention analysis; overdispersion; particle methods; state-space models; zero-inflation

Received July 2013; **revised** February 2014; **accepted** February 2014

1 Introduction

Time series of counts are frequently encountered in practice (Kedem and Fokianos, 2002; Fokianos, 2011; Cameron and Trivedi, 2013). Accordingly, the modelling and analysis of such data has received considerable attention in the literature. In principle, count time series may exhibit three distinctive features: overdispersion, zero-inflation and temporal correlation. Developing a modelling framework that is sufficiently general to accommodate all three of these characteristics poses a challenge.

In general, time series models can be classified as either observation-driven or parameter-driven (Cox, 1981). These two types of models differ in the way they account for autocorrelation. With observation-driven models, temporal correlation

Address for correspondence: Ming Yang, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, HSPH 2-435B, Boston, MA 02115, USA.
E-mail: mingyang@hsph.harvard.edu

 10.1177/1471082X14535530

between adjacent observations is directly characterized as a function of past responses. In contrast, with parameter-driven models, an unobserved latent process is employed to account for serial correlation. Conditioning on the latent process, the observations are assumed to be independently distributed.

Several modelling frameworks have been developed for count time series that attempt to account for overdispersion and temporal correlation, based on both the observation-driven approach (Zeger and Qaqish, 1988; Davis *et al.*, 2003; Freeland and McCabe, 2004; Fokianos *et al.*, 2009; Zhu, 2010) and the parameter-driven approach (Zeger, 1988; Oh and Lim, 2001; Chan and Ledolter, 1995; Nelson and Leroux, 2006; Davis and Wu, 2009). Recently, a number of frameworks have also been proposed to account for zero-inflation (Wang, 2001; Dalrymple *et al.*, 2003; Yau *et al.*, 2004; Lee *et al.*, 2005; Jazi *et al.*, 2012; Zhu, 2012; Yang *et al.*, 2013). In this article, we provide a unified approach for modelling count time series. Specifically, we propose a flexible class of state-space (dynamic) models that simultaneously accounts for autocorrelation, overdispersion and zero-inflation. Certain models that have been previously introduced may be viewed as special cases of this model class.

With count time series, the response distribution is often markedly non-Gaussian. Thus, in the state-space setting, traditional methods based on the Kalman filter and Kalman smoother cannot be used for parameter estimation. We therefore resort to Monte Carlo methods, and devise a Monte Carlo expectation-maximization (MCEM) algorithm based on the particle filter (Gordon *et al.*, 1993) and particle smoother (Godsill *et al.*, 2004). Our MCEM algorithm is an extension of the exact EM algorithm for linear and Gaussian state-space models (Shumway and Stoffer, 1982). A similar MCEM algorithm has been proposed by Kim and Stoffer (2008) to fit stochastic volatility models.

The paper is organized as follows. In Section 2, we introduce a class of dynamic models for zero-inflated count time series in a state-space framework. We outline the E-step and M-step of the MCEM algorithm in Section 3. Section 4 contains details regarding the particle methods that are used in the E-step of the MCEM algorithm. Analysis of simulated data is presented in Section 5. In Section 6, we illustrate the proposed methodology using occupational injury data initially analyzed by Yau *et al.* (2004). Section 7 concludes with a discussion of future directions.

2 Dynamic models

The Poisson distribution has been widely used to model discrete count data. In the presence of overdispersion, the negative binomial (NB) distribution is often employed as an alternative. To model count data with excess zeros, Lambert (1992) introduced a regression model based on the zero-inflated Poisson (ZIP) distribution. Zero-inflated NB (ZINB) models have also been developed (e.g., Yau *et al.*, 2003) to simultaneously account for overdispersion and zero-inflation.

Based on the ZINB distribution, we introduce a class of dynamic models for zero-inflated and overdispersed count time series. This class is formulated and developed in the state-space framework.

First, to accommodate temporal correlation in the series, we consider a stationary autoregressive process $\{z_t\}$ of order p (AR(p)) such that

$$z_t = \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} + \epsilon_t, \tag{2.1}$$

where ϵ_t is a Gaussian white noise process with mean 0 and variance σ^2. Here, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$ is a p-dimensional vector that consists of the autoregressive coefficients of $\{z_t\}$. For the AR(p) process to be stationary, it is necessary (although not sufficient) that both

$$\phi_1 + \cdots + \phi_p < 1 \text{ and } |\phi_p| < 1.$$

Conditioning on the current state z_t, the observation y_t is assumed to follow a ZINB distribution with probability mass function defined as follows:

$$f_{Y_t}(y_t | z_t; \omega, k, \lambda_t) = \omega I_{(y_t=0)} + (1-\omega)\frac{\Gamma(k+y_t)}{\Gamma(k)y_t!} p_t^k (1-p_t)^{y_t}, \tag{2.2}$$

where $p_t = k/(k + \lambda_t)$ denotes the probability of success in an NB distribution and λ_t is an intensity parameter. The intensity λ_t is explicitly linked to the latent state z_t through a log-linear model, to be specified in equation (2.4). In the preceding, $I_{(y_t=0)}$ is an indicator taking the value 1 when $y_t = 0$ and 0 otherwise. For simplicity, the dispersion parameter k (default parameterization in R) and the zero-inflation parameter ω are treated as constant, although they could be time-varying. Another popular parameterization of the dispersion parameter is to use $\tau = 1/k$ (e.g., in SAS). For the remainder of the paper, both parameterizations for the NB dispersion parameter will be used, the choice depending on which is more convenient for the purpose at hand.

The conditional mean and variance of the ZINB distribution are given by

$$E(Y_t | z_t) = \lambda_t(1-\omega) \text{ and } \mathrm{Var}(Y_t | z_t) = \lambda_t(1-\omega)(1+\omega\lambda_t+\tau\lambda_t), \tag{2.3}$$

respectively. Based on (2.3), the variance-to-mean ratio (i.e., $1+(\omega+\tau)\lambda_t$) is greater than or equal to 1, which implies that the presence of zero-inflation ($\omega > 0$) and NB dispersion ($\tau > 0$) will both contribute to the overdispersion of $Y_t | z_t$. It is worth noting that the presence of the correlated random effects $\{z_t\}$ will not only account for autocorrelation but also for overdispersion.

Similar to Poisson and NB regression models for independent data, the following log-linear model is used to characterize the intensity parameter λ_t:

$$log\lambda_t = logw_t + \mathbf{x}_t^\top \boldsymbol{\beta} + z_t, \tag{2.4}$$

where $\mathbf{x}_t$ is a set of explanatory variables and $\boldsymbol{\beta}$ is the vector regression coefficients. Here, $\log w_t$ denotes an offset variable, which can be used to account for varying population size over time.

The dynamic ZINB model defined by (2.1), (2.2) and (2.4) is very general and includes many important models as special cases. A list of the types of models subsumed by the dynamic ZINB model is provided in Table 1. For instance, the dynamic ZINB model reduces to a dynamic ZIP model when $\tau = 1/k \to 0$, to an ordinary ZINB model for independent data when $\sigma = 0$, and to an ordinary Poisson model for independent data when $\tau \to 0, \sigma = 0$ and $\omega = 0$. We emphasize that excess dispersion is not only introduced through the NB distribution ($\tau > 0$), yet also through zero-inflation ($\omega > 0$) and latent autocorrelation ($\sigma > 0$). When the data is sufficiently and suitably overdispersed, the separation of the effects of overdispersion induced by τ, ω, σ may be delineated. Otherwise, we recommend fitting a simpler model to avoid estimation problems due to weak identifiability, which results in a likelihood surface that exhibits gradual curvature in a neighborhood of the global maximum.

The NB distribution can accommodate a certain frequency of zeros. However, to characterize series where the prevalence of zeros is beyond the capacity of the NB distribution, the mixture parameter ω can assign positive weight to the degenerate distribution at zero. In general, a value of ω near zero will reflect a frequency of zeros that is compatible with the NB distribution, whereas a larger value will reflect zero-inflation. A value of ω near 1 corresponds to a series comprised of sparse positive counts with a preponderance of zeros. Such a time series could be difficult to model, since only the positive values inform the autocorrelation structure.

Because one can view the ZIP model as a two-component mixture (Lambert, 1992) and the NB model as a Poisson-gamma mixture (Lawless, 1987), it is natural to rewrite the dynamic ZINB model in the following hierarchial form:

$$
\begin{aligned}
\mathbf{s}_t \mid \mathbf{s}_{t-1} &\sim \mathcal{N}_p(\boldsymbol{\Phi}\mathbf{s}_{t-1}, \boldsymbol{\Sigma}), \\
u_t &\sim \text{Bernoulli}(\omega), \\
v_t &\sim \text{Gamma}(k, 1/k), \\
y_t \mid \mathbf{s}_t, u_t, v_t &\sim \text{Poisson}((1-u_t)v_t\lambda_t),
\end{aligned}
$$

Table 1 Types of statistical models implied by the general dynamic ZINB model.

Zero-inflation	Overdispersion	Autocorrelation	Model
No	No	No	Poisson regression
		Yes	Dynamic Poisson regression
	Yes	No	NB regression
		Yes	Dynamic NB regression
Yes	No	No	ZIP regression
		Yes	Dynamic ZIP regression
	Yes	No	ZINB regression
		Yes	Dynamic ZINB regression

Source: Authors' own.

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

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

Note that the covariance matrix $\boldsymbol{\Sigma}$ is not positive definite. This should not be surprising as only the first element of $\mathbf{s}_t$ contributes to the likelihood function. All the remaining $(p-1)$ elements of $\mathbf{s}_t$ simply serve as auxiliary variables so that the univariate AR(p) process can be written as a p-dimensional vector AR(1) (VAR(1)) process.

3 MCEM algorithm

The marginal likelihood of the observed data $y_{1:n} = (y_1, \ldots, y_n)^\top$ cannot be expressed analytically since (i) the response distribution is non-Gaussian, and (ii) the random effects $\{z_t\}$ are autocorrelated. For these reasons, direct maximization of the marginal likelihood, if possible, would be extremely difficult. Therefore, instead of using gradient-based methods (e.g., Newton–Raphson), we resort to the EM algorithm (Dempster *et al.*, 1977), which is a popular method for calculating maximum likelihood estimators (MLEs) for models involving missing data and/or unobservable latent variables.

Assuming the latent processes $\mathbf{s}_{0:n} = (\mathbf{s}_0^\top, \mathbf{s}_1^\top, \ldots, \mathbf{s}_n^\top)^\top$, $u_{1:n} = (u_1, \ldots, u_n)^\top$ and $v_{1:n} = (v_1, \ldots, v_n)^\top$ could be observed, we 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:

$$\begin{aligned} L_c(\boldsymbol{\theta}) &= \pi(\mathbf{s}_{0:n}, u_{1:n}, v_{1:n}, y_{1:n}) \\ &= \pi(\mathbf{s}_{0:n}, u_{1:n}, v_{1:n})\pi(y_{1:n} \mid \mathbf{s}_{0:n}, u_{1:n}, v_{1:n}) \\ &= \pi(\mathbf{s}_{0:n})\pi(u_{1:n})\pi(v_{1:n})\pi(y_{1:n} \mid \mathbf{s}_{0:n}, u_{1:n}, v_{1:n}) \\ &= \pi(\mathbf{s}_0)\prod_{t=1}^{n}\pi(\mathbf{s}_t \mid \mathbf{s}_{t-1})\prod_{t=1}^{n}\pi(u_t)\prod_{t=1}^{n}\pi(v_t)\prod_{t=1}^{n}\pi(y_t \mid \mathbf{s}_t, u_t, v_t). \end{aligned} \tag{3.1}$$

Here, $\boldsymbol{\theta} = (\omega, k, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \sigma)^\top$ is the vector of unknown parameters. The initial state vector $\mathbf{s}_0$ is often assumed to be normally distributed with the mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. Based on our experience, the choice of $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ has very little effect on estimated parameters. For simplicity, for the results compiled and reported for this paper, we use $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_p$.

Due to the orthogonal decomposition in (3.1), the complete data log-likelihood (up to an additive constant) is given by

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

To implement the EM algorithm, we need to compute the expectation of $l_c(\boldsymbol{\theta})$ given the observed data $y_{1:n}$. To simplify the notation, we let $\mathbf{A}_t^{(j)}$, $\mathbf{b}_t^{(j)}$, $c_t^{(j)}$, $d_t^{(j)}$, $e_t^{(j)}$, $f_t^{(j)}$ and $g_t^{(j)}$ 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, v_t, $\log v_t$ and $(1-u_t)v_t\exp(z_t)$, respectively. In the E-step of the algorithm, the conditional expectation of $l_c(\boldsymbol{\theta})$ is given by

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

Unlike the exact EM algorithm for linear and Gaussian state-space models (Shumway and Stoffer, 1982), there is no analytical form for the preceding conditional expectation due to the non-normality of the data. One approach to approximate the conditional expectation is to simulate dependent posterior samples using Markov chain Monte Carlo (MCMC). This approach has been adopted by Chan and Ledolter (1995) for fitting a Poisson state-space model. Another approach is to simulate independent posterior realizations based on the particle filter (Gordon *et al.*, 1993) and the particle smoother (Godsill *et al.*, 2004). This approach has been employed by Kim and Stoffer (2008) for fitting stochastic volatility models in the presence of irregular sampling. Since it is more efficient to use independent realizations in an MCEM algorithm (Levine and Casella, 2001), we follow the latter approach (Kim and Stoffer, 2008) and approximate the conditional expectation via particle methods. The details of the particle methods for the dynamic ZINB model will be presented in the next section.

After the Monte Carlo E-step, the M-step of the algorithm becomes straightforward due to the orthogonal decomposition of the complete data log-likelihood. In the M-step, the following partial derivatives are applied to maximize $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(j)})$:

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

where $\Psi_0(k) = \frac{d}{dk}\log\Gamma(k)$ is the digamma function. At the jth iteration, closed-form solutions exist to update $\omega^{(j+1)}$, $\boldsymbol{\phi}^{(j+1)}$ and $\sigma^{(j+1)}$:

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

In addition, we can easily obtain $k^{(j+1)}$ and $\boldsymbol{\beta}^{(j+1)}$ through iterative algorithms. For example, updating $\boldsymbol{\beta}^{(j+1)}$ is equivalent to fitting a weighted Poisson regression.

Once we obtain the MLE through the MCEM algorithm, we apply Louis's formula (Louis, 1982) to compute the observed information matrix $\mathbf{I}_o(\boldsymbol{\theta})$. According to the missing information principle, the complete data information is the sum of the observed data information and the missing data information. Therefore, we have

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

where the complete information matrix $\mathbf{I}_c(\boldsymbol{\theta})$ and the missing data information matrix $\mathbf{I}_m(\boldsymbol{\theta})$ are defined as follows:

$$\mathbf{I}_c(\boldsymbol{\theta}) = \mathrm{E}\left(-\frac{\partial^2 l_c}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \mid y_{1:n}\right), \tag{3.2}$$

$$\mathbf{I}_m(\boldsymbol{\theta}) = \mathrm{E}\left(\frac{\partial l_c}{\partial\boldsymbol{\theta}}\frac{\partial l_c}{\partial\boldsymbol{\theta}^\top} \mid y_{1:n}\right) - \mathrm{E}\left(\frac{\partial l_c}{\partial\boldsymbol{\theta}} \mid y_{1:n}\right)\mathrm{E}\left(\frac{\partial l_c}{\partial\boldsymbol{\theta}^\top} \mid y_{1:n}\right). \tag{3.3}$$

The elements of $\partial l_c/\partial\boldsymbol{\theta}$ and $\partial^2 l_c/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ are given in the Appendix. Again, the particle methods can be used to approximate the conditional expectations in (3.2) and (3.3). The variance–covariance matrix is then obtained by taking the inverse of the observed information matrix.

4 Particle methods

In this section, we present the particle methods that are used to approximate the conditional expectations in the Monte Carlo E-step. Particle filtering (Gordon *et al.*, 1993) and particle smoothing (Godsill *et al.*, 2004) belong to the class of sequential Monte Carlo (SMC) methods (Doucet *et al.*, 2001). 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 subsequently defined) 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 and Stoffer (2008, pp. 828–829).

For the dynamic ZINB model, we implement particle filtering by first generating $\mathbf{s}_{0|0}^{(i)} \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. Then for $t = 1, \ldots, n$, we complete the following steps to produce a set of N particles ($i = 1, \ldots, N$) at each time point t.

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

(F.2) Compute the filtering weights

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

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

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

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

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

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 $\left(\mathbf{s}_{n|n}^{(r)}, u_{n|n}^{(r)}, v_{n|n}^{(r)}\right) = \left(\mathbf{s}_{n|n}^{(i)}, u_{n|n}^{(i)}, v_{n|n}^{(i)}\right)$ with probability $q_{n|n-1}^{(i)}$. Then for $t = n - 1, \ldots, 1$:

(S.1) Calculate the smoothing weights

$$q_{t|n}^{(i)} \propto q_{t|t-1}^{(i)} \exp\left\{-\frac{1}{2\sigma^2}\left(z_{t+1|n}^{(i)} - \boldsymbol{\phi}^{\mathrm{T}}\mathbf{s}_{t|t}^{(i)}\right)^2\right\}$$
$$\times\left\{\omega^{u_{t+1|n}^{(i)}}(1-\omega)^{1-u_{t+1|n}^{(i)}}\right\}\times\left\{\left(v_{t+1|n}^{(i)}\right)^{k-1}\exp\left(-kv_{t+1|n}^{(i)}\right)\right\}.$$

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

We obtain *independent* realizations by repeating the preceding steps for $r = 1, \ldots, R$. The forward-filtering and backward-smoothing procedures are the non-linear and non-Gaussian extensions of the Kalman filter and the Kalman smoother.

5 Simulation

5.1 Model fitting based on simulated data

To evaluate the performance of the MCEM algorithm, we simulate a count series of length 200 from a dynamic ZINB + AR(2) model with the following linear predictor:

$$\log \lambda_t = \beta_0 + \beta_1 x_t + z_t,$$

where $x_t = I_{(t>100)}$ is a dummy variable indicating whether the time index t is greater than 100. In applications involving intervention analysis, often referred to as interrupted time series analyses, this variable will represent whether the time is before or after the intervention. The parameters of the true model are $\omega = 0.3$, $\tau = 0.4$, $\beta_0 = 2$, $\beta_1 = -1$, $\phi_1 = 0.8$, $\phi_2 = -0.6$ and $\sigma = 0.5$. The autoregressive coefficients (i.e., ϕ_1 and ϕ_2) are chosen such that the AR(2) process is stationary.

A total of eight dynamic models are fit to the simulated series: Poisson + AR(1), NB + AR(1), ZIP + AR(1), ZINB + AR(1), Poisson + AR(2), NB + AR(2), ZIP + AR(2) and ZINB + AR(2). When fitting these dynamic models, the number of particles (N) and the number of replications (R) are chosen to be 500. The MCEM algorithm is stopped after 500 iterations. Table 2 shows the parameter estimates and their associated standard errors for the eight candidate models.

Table 2 Parameter estimates (standard errors) for eight dynamic models fit to data simulated from a dynamic ZINB + AR(2) model with true parameters $\omega = 0.3$, $\tau = 0.4$, $\beta_0 = 2$, $\beta_1 = -1$, $\phi_1 = 0.8$, $\phi_2 = -0.6$ and $\sigma = 0.5$.

	Poisson + AR(1)	NB + AR(1)	ZIP + AR(1)	ZINB + AR(1)
ω			0.266 (0.043)	0.230 (0.052)
τ		2.044 (0.288)		0.859 (0.325)
β_0	1.152 (0.132)	2.075 (0.113)	1.981 (0.134)	2.254 (0.123)
β_1	–1.349 (0.200)	–1.340 (0.173)	–1.385 (0.215)	–1.304 (0.197)
ϕ_1	0.032 (0.091)	0.201 (0.538)	0.187 (0.126)	0.311 (0.229)
ϕ_2				
σ	1.527 (0.128)	0.204 (0.106)	0.964 (0.102)	0.408 (0.203)

	Poisson + AR(2)	NB + AR(2)	ZIP + AR(2)	ZINB + AR(2)
ω			0.265 (0.042)	0.267 (0.048)
τ		1.726 (0.294)		0.278 (0.431)
β_0	1.131 (0.708)	1.938 (0.203)	1.969 (0.131)	2.083 (0.140)
β_1	–1.355 (0.380)	–1.347 (0.239)	–1.370 (0.241)	–1.384 (0.425)
ϕ_1	0.048 (0.092)	0.655 (0.156)	0.370 (0.125)	0.687 (0.189)
ϕ_2	–0.112 (0.093)	–0.781 (0.303)	–0.399 (0.124)	–0.696 (0.280)
σ	1.524 (0.286)	0.337 (0.479)	0.866 (0.094)	0.529 (0.217)

Source: Authors' own.

Figure 1 displays trace plots of the log-likelihood for the eight candidate models. Unlike the exact EM algorithm for linear and Gaussian state-space models (Shumway and Stoffer, 1982), the log-likelihood in the MCEM algorithm is not guaranteed to increase at each iteration. Instead, as illustrated in the figure, the log-likelihood will increase dramatically in the first several iterations, and then stabilize as the estimated parameters approach the MLE.

Trace plots illustrating scaled changes for the parameter estimates are presented in Figure 2. We note that some of the parameters are still changing substantially even when the log-likelihood has plateaued, especially for the NB + AR(1) and NB + AR(2) models. Therefore, for the MCEM algorithm, a stopping rule solely based on log-likelihood change is unreliable. We recommend always checking the trace plots of parameter estimates for stabilization.

5.2 Simulation study

To investigate the finite sample distributional properties of the parameter estimators from the MCEM algorithm, an extensive simulation study is warranted. However, due to the computational requirements of the algorithm, a large scale study is prohibitively time consuming. As an alternative, we have compiled results for a small scale study

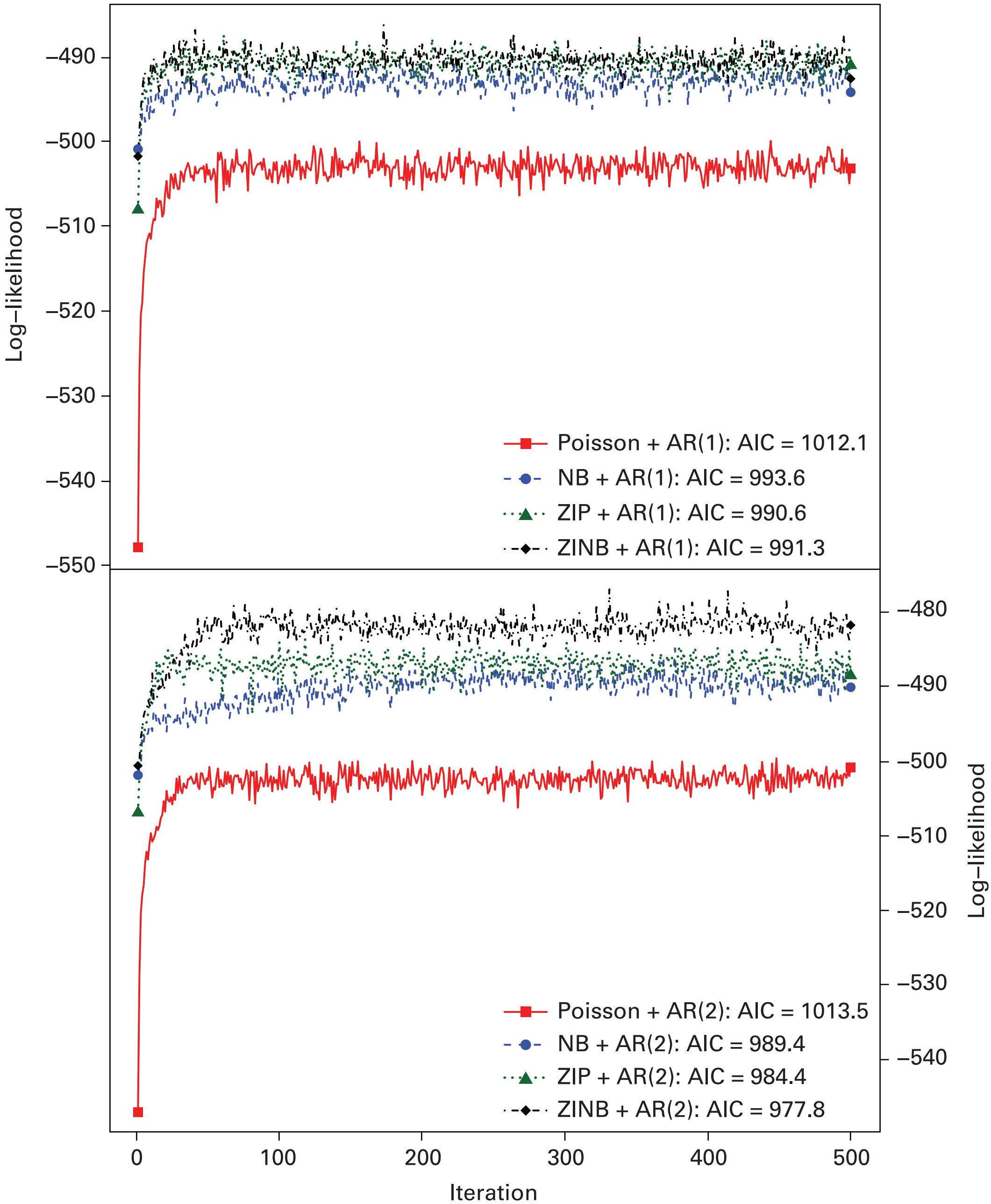

Figure 1 Trace plots of the log-likelihood for eight dynamic models fit to data simulated from a dynamic ZINB + AR(2) model.

Source: Authors' own.

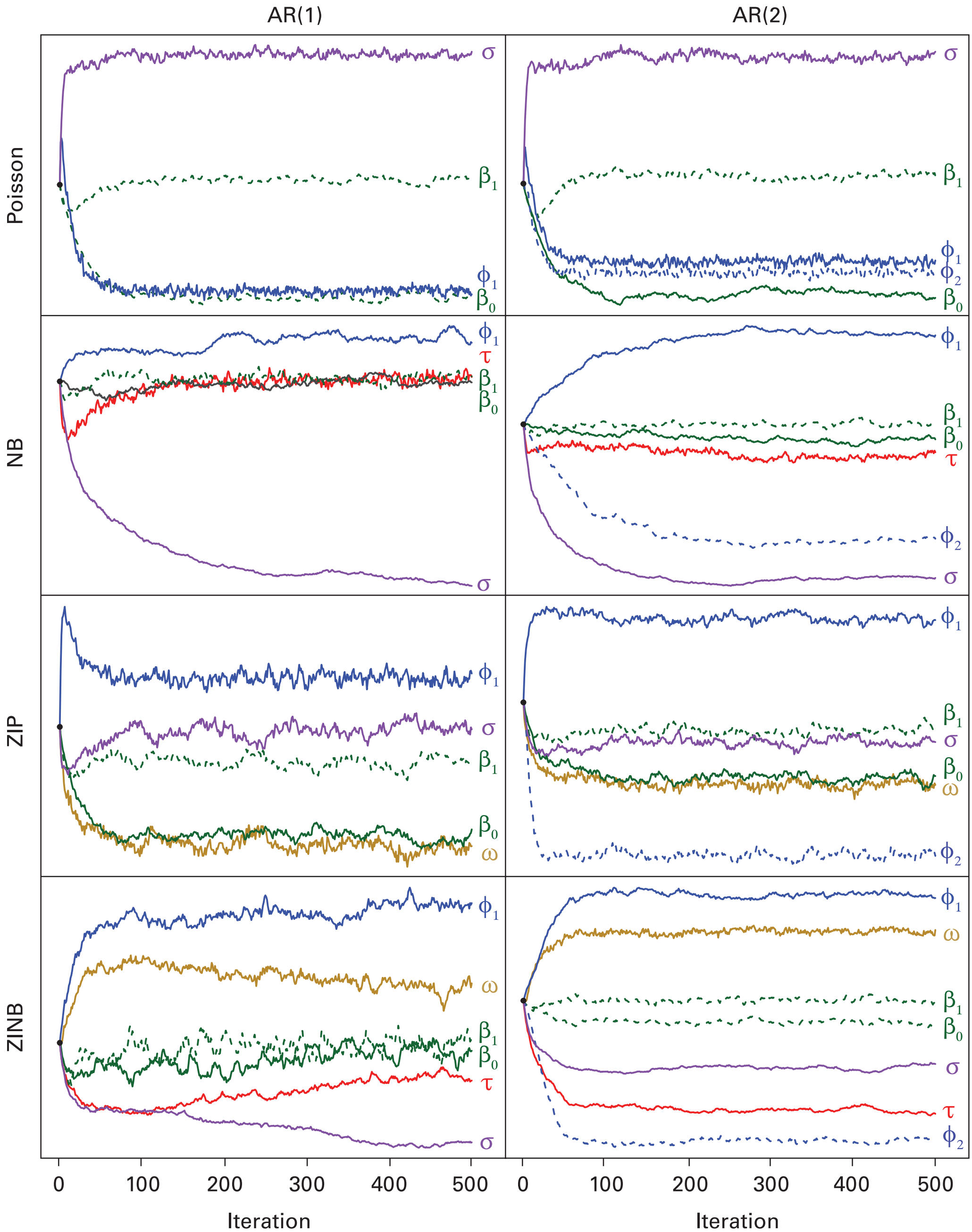

Figure 2 Trace plots of scaled changes in parameter estimates from starting values. Estimates are for eight dynamic models fit to data simulated from a dynamic ZINB + AR(2) model.

Source: Authors' own.

to examine the sampling distributions for estimators based on the eight model structures considered in the previous subsection. For each model structure, 100 series of length 200 are generated, the MCEM algorithm is used to fit models based on these series and the parameter estimates and standard errors are recorded. To reduce time requirements, we set $N = 100$ and $R = 100$, and we stopped the MCEM algorithm after 200 iterations. For the sake of brevity, we present a comprehensive summary of the results based on the dynamic ZIP + AR(2) model, and comment on the results not featured. We use the parameter values corresponding to the data generating model in the previous subsection, with the exception of τ, which is set to zero.

To assess approximate normality of the estimators, Q-Q plots based on the sets of replicated estimates are presented in Figure 3. In constructing these plots, two sets of estimates have been omitted due to convergence failures with the fitted model. These plots illustrate that approximate normality holds for the sampling distributions of the estimators. However, the majority of the plots reflect slightly longer tails in the empirical distribution of the replicates than the tails for the reference normal curve. This behaviour should be attenuated as the sample size is increased. Table 3 provides the true parameter values, along with the means and medians of the estimates, the empirical standard deviations (ESD) of the estimates and the means of the asymptotic standard errors (ASE). We note a minor degree of bias associated with the estimation of the AR coefficients. In general, the ESDs are reasonably close to the average ASE. Therefore, it seems appropriate to calculate the standard errors based on equations (3.2)–(3.3).

The results for the sets based on the dynamic ZIP and Poisson models are similarly promising. The results for the sets based on the ZINB and NB models exhibit less adherence to the ideal asymptotic properties. This is not surprising, due to the problems with weak identifiability previously mentioned.

6 Application

In this section, we use the occupational injury data featured in Yau *et al.* (2004) to illustrate our proposed methodology. The application concerns the assessment

Table 3 Summary statistics for replicated parameter estimates from fitted dynamic ZIP + AR(2) models.

	True	Mean	Median	ESD	ASE
ω	0.3	0.297	0.298	0.040	0.038
β_0	2.0	2.103	2.064	0.222	0.159
β_1	–1.0	–1.067	–1.058	0.264	0.242
ϕ_1	0.8	0.710	0.704	0.127	0.173
ϕ_2	–0.6	–0.518	–0.541	0.145	0.162
σ	0.5	0.524	0.518	0.073	0.102

Source: Authors’ own.

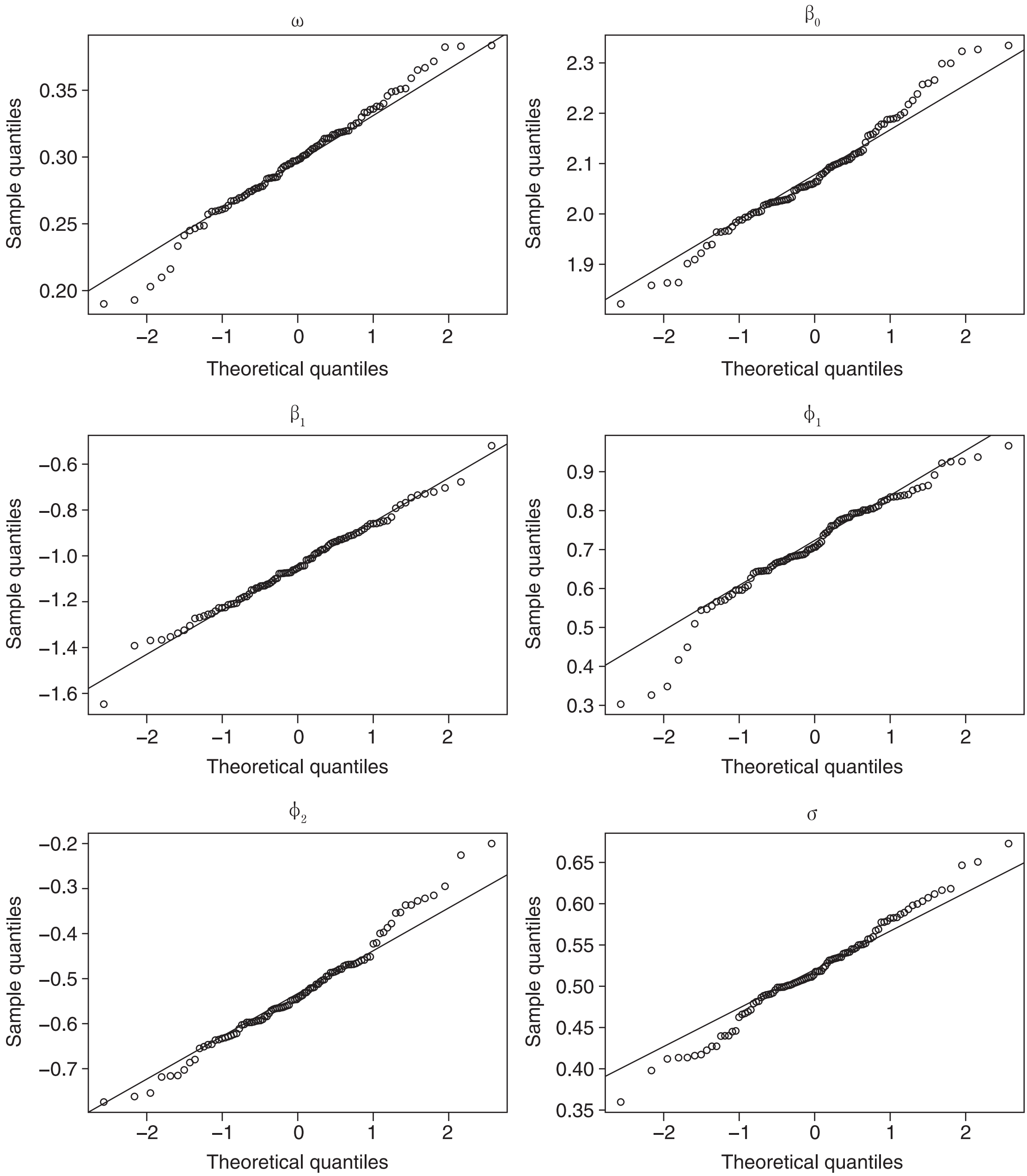

Figure 3 Q-Q plots for the estimated parameters from a dynamic ZIP + AR(2) model with true parameters $\omega = 0.3$, $\beta_0 = 2$, $\beta_1 = -1$, $\phi_1 = 0.8$, $\phi_2 = -0.6$ and $\sigma = 0.5$.

Source: Authors' own.

of a participatory ergonomics intervention in reducing the incidence of workplace injuries among a group of hospital cleaners. The data consists of monthly (4-week) counts of work-related injuries that are routinely reported at an aggregate population level from July 1988 to October 1995. A participatory ergonomics intervention was introduced in the middle of the study in November 1992. During the study period, a large number of zero counts are observed due to the heterogeneity in risk and the dynamic worker population (Yau *et al*., 2004). Since the injury count series contain excess zeros relative to a Poisson distribution, Yau *et al.* (2004) modelled the data using a ZIP mixed autoregression, which is a special case of the general dynamic ZINB model proposed in this paper.

Our analyses of the injury count series focus on two objectives. First, we wish to investigate whether there is additional autocorrelation that is not explained by the simple AR(1) structure. Second, we wish to investigate the presence of unexplained excess dispersion, since any overdispersion in the counts may not be fully accounted for by the correlated random effects $\{z_t\}$. To address these issues, we fit eight candidate models with the same structures as the models considered in Section 5. Specifically, we employ the linear predictor

$$\log \lambda_t = \beta_0 + \beta_1 x_t + z_t, \qquad t = 1, \ldots, 96,$$

where $x_t = I_{(t>57)}$ is a dummy variable indicating whether the time index t is greater than the intervention time (57 months). Thus, β_1 reflects the reduction in injury risk due to the intervention. We consider the dynamic model structures Poisson + AR(1), NB + AR(1), ZIP + AR(1), ZINB + AR(1), Poisson + AR(2), NB + AR(2), ZIP + AR(2) and ZINB + AR(2). The Akaike (1974) information criterion (AIC) is used to guide the selection of an optimal model.

Table 4 features results for the four candidate models with the AR(1) correlation structure. All four models indicate a significant reduction of work-related injuries after the introduction of the intervention ($\beta_1 < 0$). The standard errors of $\widehat{\beta}_1$ are very similar across the four models, but the magnitudes of $\widehat{\beta}_1$ are less pronounced

Table 4 Parameter estimates (standard errors) for dynamic models fit to the injury count series.

	Poisson + AR(1) (AIC = 316.0)	NB + AR(1) (AIC = 316.2)	ZIP + AR(1) (AIC = 308.6)	ZINB + AR(1) (AIC = 311.3)
ω			0.304 (0.084)	0.312 (0.084)
τ		0.792 (0.303)		0.042 (0.055)
β_0	0.331 (0.213)	0.633 (0.184)	0.852 (0.208)	0.896 (0.185)
β_1	–1.124 (0.347)	–1.086 (0.308)	–0.905 (0.347)	–0.878 (0.365)
ϕ	0.293 (0.202)	0.574 (0.302)	0.520 (0.369)	0.576 (0.424)
σ	0.841 (0.155)	0.243 (0.176)	0.403 (0.202)	0.342 (0.172)

Source: Authors' own.

after accounting for zero-inflation. The ZIP + AR(1) model that accounts for zero-inflation has the smallest AIC. The more complicated ZINB + AR(1) model does not outperform the ZIP + AR(1) model, indicating there is no need to incorporate a latent gamma component to explain the overdispersion.

Figure 4 displays the particle smoothing of the series for the four candidate models with the AR(1) correlation structure. Not surprisingly, the fitted values from the ZIP

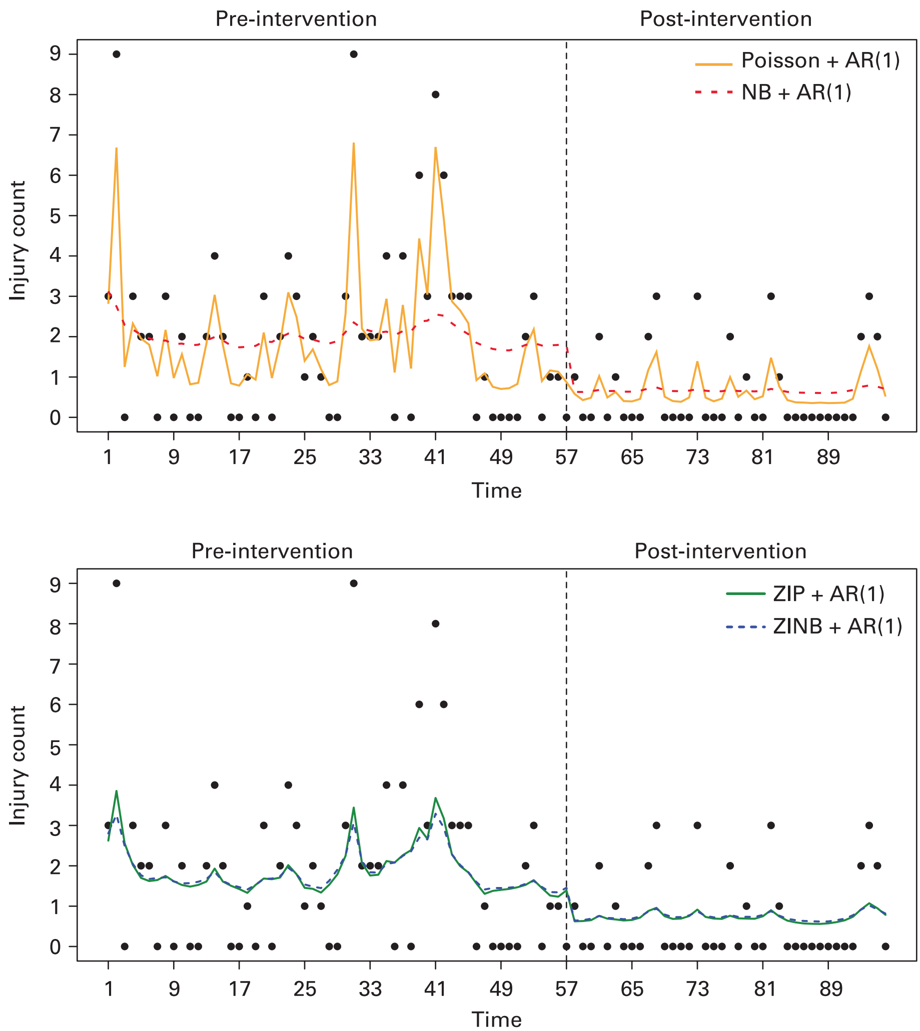

Figure 4 Particle smoothing of the state based on dynamic models fit to the injury count series. Dots represent the observed counts.

Source: Authors' own.

+ AR(1) and ZINB + AR(1) models are very similar. Failure to account for zero-inflation in the Poisson + AR(1) and NB + AR(1) models results in characteristically different fitted values.

The four models featuring the AR(2) correlation structure do not fit the series appreciably better than their AR(1) counterparts, and based on AIC, the added complexity of the AR(2) structure is unwarranted. Thus, our findings here reinforce the use of the dynamic ZIP + AR(1) model proposed by Yau *et al.* (2004) for the analysis of this data. However, our modelling framework is very general in the sense that all four count data distributions (Poisson, NB, ZIP, ZINB) and any order of autoregressive correlation can be examined.

7 Discussion

Count time series with excess zeros are encountered in many areas of statistical application. Although the Poisson and NB distributions have been widely used in practice to model 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.

In this article, we propose a class of state-space models to analyze time series of counts containing excess zeros. Both simulated and real examples are presented to illustrate the proposed methodology. To implement the modelling procedures introduced in this paper and in Yang *et al.* (2013), we have developed an R package called ZIM (zero-inflated models). This package is currently available on CRAN (Comprehensive R Archive Network). Due to the possibility of estimation problems caused by weak identifiability, we do not recommend fitting a dynamic ZINB model when sample information is limited. Specifically, if only a short sequence of observations is available (e.g., $n < 60$), or the series is comprised of very sparse counts relative to high zero-inflation, the complexity of the dynamic ZINB model may lead to poor results.

The implementation of an effective, general stopping criterion for the MCEM algorithm presents a daunting challenge, especially when weak identifiability is a potential concern. Many stopping criteria commonly used in the literature are appealing. For example, one can terminate the algorithm when the log-likelihood starts to stabilize, or when the score vector is close to zero. Our experience is that these rules fall short in the presence of weak identifiability. We have found some success in building a stopping criterion using an iteration threshold in conjunction with log-likelihood trace plots. Specifically, we terminate the algorithm after 500 iterations, and examine the resulting trace plots. The development of a defensible, automated stopping rule is a worthwhile practical objective that we hope to accomplish in our future research.

For the methodological developments presented in this paper, the parameters ω and k are assumed to be constant over time. With some applications, it may be desirable to allow these parameters to evolve over time. In such situations, one could formulate logistic regression and log-linear regression models to impose structure on ω_t and k_t, respectively, which would allow for the incorporation of potentially relevant predictors (e.g., a pre-/post-intervention indicator). However, non-identifiability could arise as the complexity of the model increases. Further research on this issue is needed.

With the advancement of modern computing, Bayesian methods have become increasingly popular in applied statistics. In future work, we hope to develop a fully Bayesian modelling framework for zero-inflated time series using either MCMC algorithms (e.g., Gibbs and Metropolis–Hastings sampling) or the integrated nested Laplace approximations (INLA) approach proposed by Rue *et al.* (2009). Specifically, we plan to develop an efficient particle MCMC algorithm (Andrieu *et al.*, 2010) for the dynamic models presented in this article. Such a hybrid algorithm will allow one to efficiently sample from high-dimensional probability distributions. We will compare the (particle) MCMC approach to the approximate INLA approach.

Acknowledgements

The authors wish to extend their appreciation to two anonymous reviewers for carefully reading the original version of this manuscript, and for providing constructive suggestions that served to improve the presentation and content.

Appendix

This appendix contains the first and second partial derivatives of the complete data log-likelihood for the general dynamic ZINB model. The elements of $\partial l_c / \partial \boldsymbol{\theta}$ are given by

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

$$\frac{\partial l_c}{\partial k} = n\{1 + \log k - \Psi_0(k)\} + \sum_{t=1}^{n}(\log v_t - v_t),$$

$$\frac{\partial l_c}{\partial \boldsymbol{\beta}} = \sum_{t=1}^{n}(1-u_t)\{y_t - v_t w_t \exp(\mathbf{x}_t^\top \boldsymbol{\beta} + z_t)\}\mathbf{x}_t,$$

$$\frac{\partial l_c}{\boldsymbol{\phi}} = \frac{1}{\sigma^2}\sum_{t=1}^{n}(z_t - \boldsymbol{\phi}^\top \mathbf{s}_{t-1})\mathbf{s}_{t-1},$$

$$\frac{\partial l_c}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{t=1}^{n}(z_t - \boldsymbol{\phi}^\top \mathbf{s}_{t-1})^2.$$

Due to the orthogonal decomposition of the complete data likelihood, most of the off-diagonal elements in $\partial^2 l_c/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ are zeros. The following are the diagonal elements and the non-zero off-diagonal elements:

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

$$\frac{\partial^2 l_c}{\partial k\partial k} = n\left\{\frac{1}{k} - \Psi_1(k)\right\},$$

$$\frac{\partial^2 l_c}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = -\frac{1}{\sigma^2}\sum_{t=1}^{n}(1-u_t)v_t w_t \exp(\mathbf{x}_t^\top\boldsymbol{\beta} + z_t)\mathbf{x}_t\mathbf{x}_t^\top,$$

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

$$\frac{\partial^2 l_c}{\partial\sigma\partial\sigma} = \frac{n}{\sigma^3} - \frac{3}{\sigma^4}\sum_{t=1}^{n}(z_t - \boldsymbol{\phi}^\top\mathbf{s}_{t-1})^2,$$

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

Here, $\Psi_0(k) = \frac{d}{dk}\log\Gamma(k)$ and $\Psi_1 = \frac{d}{dk}\Psi_0(k)$ are the digamma and trigamma functions, respectively.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–23.

Andrieu C, Doucet A and Holenstein R (2010) Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B*, **72**, 269–342.

Cameron AC and Trivedi PK (2013) *Regression analysis of count data*. 2nd edn. Cambridge University Press.

Chan KS and Ledolter J (1995) Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, **90**, 242–52.

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

Dalrymple ML, Hudson IL and Ford RPK (2003) Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Computational Statistics & Data Analysis*, **41**, 491–504.

Davis RA, Dunsmuir WTM and Streett SB (2003) Observation-driven models for Poisson counts. *Biometrika*, **90**, 777–90.

Davis RA and Wu R (2009) A negative binomial model for time series of counts. *Biometrika*, **96**, 735–49.

Dempster AP, Laird NM and Rubin DB (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, **39**, 1–39.

Doucet A, Freitas ND and Gordon N (2001) *Sequential Monte Carlo methods in practice*. New York: Springer.

Fokianos K (2011) Some recent progress in count time series. *Statistics*, **45**, 49–58.

Fokianos K, Rahbek A and Tjøstheim D (2009) Poisson autoregression. *Journal of the American Statistical Association*, **104**, 1430–9.

Freeland RK and McCabe BPM (2004) Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, **25**, 701–22.

Godsill SJ, Doucet A and West M (2004) Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, **99**, 156–68.

Gordon NJ, Salmond DJ and Smith AFM (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F, Radar and Signal Processing*, **140**, 107–13.

Jazi MA, Jones G and Lai CD (2012) First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis*, **33**, 954–63.

Kedem B and Fokianos K (2002) *Regression models for time series analysis*. New Jersey: Wiley.

Kim J and Stoffer DS (2008) Fitting stochastic volatility models in the presence of irregular sampling via particle methods and the EM algorithm. *Journal of Time Series Analysis*, **29**, 811–33.

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

Lawless JF (1987) Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, **15**, 209–25.

Lee AH, Wang K, Yau KKW, Carrivick PJW and Stevenson MR (2005) Modelling bivariate count series with excess zeros. *Mathematical Biosciences*, **196**, 226–37.

Levine RA and Casella G (2001) Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics*, **10**, 422–39.

Louis TA (1982) Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B*, **44**, 226–33.

Nelson KP and Leroux BG (2006) Statistical models for autocorrelated count data. *Statistics in Medicine*, **25**, 1413–30.

Oh MS and Lim YB (2001) Bayesian analysis of time series Poisson data. *Journal of Applied Statistics*, **28**, 259–71.

Rue H, Martino S and Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society Series B*, **71**, 319–92.

Shumway RH and Stoffer DS (1982) An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, **3**, 253–64.

Wang P (2001) Markov zero-inflated Poisson regression models for a time series of counts with excess *zeros*. *Journal of Applied Statistics*, **28**, 623–32.

Yang M, Zamba GKD and Cavanaugh JE (2013) Markov regression models for count time series with excess zeros: A partial likelihood approach. *Statistical Methodology*, **14**, 26–38.

Yau KKW, Lee AH and Carrivick PJW (2004) Modeling zero-inflated count series with application to occupational health. *Computer Methods and Programs in Biomedicine*, **74**, 47–52.

Yau KKW, Wang K and Lee AH (2003) Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, **45**, 437–52.

Zeger SL (1988) A regression model for time series of counts. *Biometrika*, **75**, 621–9.

Zeger SL and Qaqish B (1988) Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, **44**, 1019–31.

Zhu F (2010) A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, **32**, 54–67.

Zhu F (2012) Zero-inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference*, **142**, 826–39.