Time Series of Multivariate Zero-inflated Poisson Counts

Chen Zhang[†], Nan Chen[†], Linmiao Zhang[‡]
[†]Department of Industrial and Systems Engineering, National University of Singapore, Singapore
[‡]F10 Data Science, Micron Technology, Singapore

Abstract - **This paper proposes a state space model to describe multivariate autocorrelated zero-inflated count series. The model extends the classical zero-inflated Poisson distribution into multivariate cases but is able to impose different zero inflations on different dimensions. Combing the zero inflation with the log-normal mixture of independent Poisson distribution, this model allows for flexible cross-correlations of multiple counts. Furthermore, By considering the zero-inflation parameters as well as the Poisson mean parameters as latent variables evolving according to state space models, the model can capture the autocorrelations of count data. Monte Carlo EM algorithm together with particle filtering method provides a satisfactory estimation for the model parameters and the latent process.**

Keywords - **Zero-inflation Poisson (ZIP), multivariate Poisson, time series, state space model, particle filtering, Monte Carlo expectation maximization (MCEM)**

I. Introduction

Count data with many zeros are common in many applications, such as ecology, epidemiology, economics, manufacturing utilities, etc. Failure to consider the extra zeros may result in biased parameter estimations and misleading inferences. For handling univariate zero-inflated count data, [1] introduced the zero-inflated Poisson (ZIP) model, i.e.,

$$Y \sim \begin{cases} 0 \quad \text{with probability } p, \\ \text{Poisson}(\lambda), \text{with probability } 1-p \end{cases} \tag{1}$$

Later lots of variants are introduced, such as [2], [3], [4]. However, usually multiple counts observed together exhibit certain correlation with each other. For example, when manufacturing involves different types of defects, the univariate ZIP distributions are no longer appropriate. In addition, many multivariate counts are evolving over time data and also exhibit autocorrelation i.e., have serial correlation with their previous observations. For example, the number of defects in neighbour samples in manufacturing may be driven by certain common inertial elements when the sampling interval is small.

However, so far to our best knowledge, there is little literature focusing on modelling multivariate autocorrelated count data with excessive zeros. [5] first extended the multivariate Poisson distribution of [6] into the zero-inflated model. Specifically, consider a p-dimensional count process which is not zero-inflated, for every dimension, the count data X_i is represented as

$$X_i = U_i + U_0, i = 1, \ldots, p \tag{2}$$

where $U_1, \ldots, U_p, U_0$ are $p+1$ independent Poisson random variables with respective means as $\lambda_1, \ldots, \lambda p, \lambda_0$. Then for the p-dimensional zero-inflated count process $\mathbf{Y} = [Y_1, \ldots, Y_p]$, [5] considers with probability p_i, Y_i follows a Poisson distribution and all the other $Y_j, j \neq i$ equals to 0. With probability $1 - \sum_{i=1}^{p} p_i$, $\mathbf{Y}$ follows the multivariate Poisson distribution of Equation (2). However, since the model of Equation (2) assumes all the p counts have the same correlation which is introduced by U_0, its ZIP version also inherits this too restricted assumption and therefore does not allow flexible cross-correlation structure of multiple counts. Later [7], [8] and [9] introduced several multivariate ZIP distributions by extending the univariate ZIP model to multivariate cases. Specifically, they assume with probability p, the p-dimensional counts equal to $\mathbf{0}$, and with probability $1-p$, the counts follow either independent Poission distributions [8] or joint multivariate log-normal Poission distributions [7], or the generalied Poisson distribution [9]. However, these models assume the zero-inflation occurs in all the dimensions simultaneously, which is not common in practice. Furthermore, all of the models above do not consider the autocorrelation property of count data.

For univariate autocorrelated ZIP model, [10] considered the first order integer valued autoregressive model (INAR) of [11] with zero-inflated Poisson innovations. [12] later proposed a first order mixed INAR model with zero-inflated generalized power series innovations. However, these two models inherit the disadvantage of the INAR model that it can only handle positive autocorrelation. [13] proposed a nonlinear time series model by assuming $\log(\lambda)$ and logit(p) of Equation (1) follow piecewise quadratic polynomials related with time. However, so far to our best knowledge, there is no counterpart of these models for multivariate cases.

To fill this research gap, illustrated by the pioneer work, this paper proposes an easy-to-interpret state space model to describe multivariate autocorrelated zero-inflated count series. The model allows for a flexible zero inflation, cross-correlation and autocorrelation structure. To be more specific, this ZIP model is built upon the infinite mixture multivariate Poisson distribution of [14], and allows for serial dependence by considering the zero inflation parameters as well as the Poisson mean vector as latent variables evolving according to state space models. In this way the model can impose flexible zero inflation on different dimensions. This model can also allow for flexible cross-correlation as well as temporal correlation. Since the marginal unconditional likelihood function has no close form, we needs numerical integration methods for model estimation. Here we use Monte Carlo Expectation Maximisation (MCEM) algorithm, where the MC part is done by particle filtering and smoothing. Particle filtering also provides asymptotically unbiased estimation of the latent processes in a sequential way with a small computation complexity.

978-1-5090-3665-3/16/$31.00 ©2016 IEEE

The remainder of the paper is organized as follows. Section II-A introduces our proposed multivariate state space Poisson model in detail. Section III discusses the model estimation procedure. Section IV demonstrates the proposal using some numerical studies. Finally Section V concludes this paper with remarks.

II. Methodology

A. A State Space Model For Zero-inflated Autocorrelated Multivariate Counts

Consider d-dimensional count variables $\mathbf{Y}_t = [Y_{t1}, \ldots, Y_{td}]$ for time $t = 1, \cdots, T$. For each dimension at time t, let with probability ξ_{ti}, Y_{ti} follows the first process (i.e., Y_{ti} is an excess 0), and with probability $1 - \xi_{ti}$, Y_{ti} follows the second Poisson process with mean equal to λ_{ti}. Then

$$Y_{ti}|\xi_{ti}, \lambda_{ti} = \begin{cases} 0 \quad \text{with probability } \xi_{ti}, \\ \text{Poisson}(\lambda_{ti}), \text{with probability } 1 - \xi_{ti} \end{cases} \tag{3}$$

for $i = 1, \ldots, d, t = 1, \ldots, T$. Then the joint conditional distribution of $\mathbf{Y}_t$ is

$$p(\mathbf{Y}_t|\xi_t, \boldsymbol{\lambda}_t) = \prod_{i=1}^{d} p(Y_{ti}|\xi_{ti}, \lambda_{ti}), t = 1, \ldots, T. \tag{4}$$

where $\boldsymbol{\xi}_t = [\xi_{t1}, \ldots, \xi_{td}]$ and $\boldsymbol{\lambda}_t = [\lambda_{t1}, \ldots, \lambda_{td}]$. We further assume $\mathbf{Z}_t = \text{logit}(\boldsymbol{\xi}_t) = \left[\log(\frac{\xi_{ti}}{1-\xi_{ti}}), \ldots, \log(\frac{\xi_{td}}{1-\xi_{td}})\right]$, and $\mathbf{X}_t = \log(\boldsymbol{\lambda}_t) = [\log(\lambda_{t1}), \ldots, \log(\lambda_{td})]$ as independent latent random variables, i.e., $\mathbf{Z}_t \perp\!\!\!\perp \mathbf{X}_t$ following multivariate normal distribution. It is $\mathbf{Z}_t$ and $\mathbf{X}_t$ that introduce both cross-correlations and autocorrelations into $\mathbf{Y}_t$. Specifically, we consider $\mathbf{Z}_t$ and $\mathbf{X}_t$ evolve according to the state space models separately as

$$p_{\boldsymbol{\Theta}}(\mathbf{Z}_t|\mathbf{Z}_{t-1}) : \mathbf{X}_t - \boldsymbol{\mu}_Z = \boldsymbol{\Phi}_Z \cdot (\mathbf{Z}_{t-1} - \boldsymbol{\mu}_Z) + \mathbf{s}_t, \tag{5}$$

$$p_{\boldsymbol{\Theta}}(\mathbf{X}_t|\mathbf{X}_{t-1}) : \mathbf{X}_t - \boldsymbol{\mu}_X = \boldsymbol{\Phi}_X \cdot (\mathbf{X}_{t-1} - \boldsymbol{\mu}_X) + \boldsymbol{\epsilon}_t, \tag{6}$$

where $\mathbf{s}_t$ and $\boldsymbol{\epsilon}_t$ are the white noise following d-dimensional multivariate normal distributions with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_Z$ and $\boldsymbol{\Sigma}_X$ respectively, i.e., $\mathbf{s}_t \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_Z)$ and $\boldsymbol{\epsilon}_t \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_X)$. So far we have introduced all the model parameters $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_Z, \boldsymbol{\mu}_X, \boldsymbol{\Phi}_Z, \boldsymbol{\Phi}_X, \boldsymbol{\Sigma}_Z, \boldsymbol{\Sigma}_X\}$.

As long as $\boldsymbol{\Phi}_Z$ and $\boldsymbol{\Phi}_X$ satisfy

$$\det(\mathbf{I} - z\boldsymbol{\Phi}_Z) \neq 0, \quad \text{for all } |z| \leq 1, \ z \in \mathcal{C},$$
$$\det(\mathbf{I} - z\boldsymbol{\Phi}_X) \neq 0, \quad \text{for all } |z| \leq 1, \ z \in \mathcal{C},$$

$\mathbf{Z}_t$ and $\mathbf{X}_t$ are stationary, and their marginal distributions are multivariate normal with mean vector $\boldsymbol{\mu}_Z$ and $\boldsymbol{\mu}_X$, and covariance matrix $\boldsymbol{\Gamma}_Z$ and $\boldsymbol{\Gamma}_X$, where $\boldsymbol{\Gamma}_Z$ and $\boldsymbol{\Gamma}_X$ are the solutions of $\boldsymbol{\Gamma}_Z = \boldsymbol{\Phi}_Z\boldsymbol{\Gamma}_Z\boldsymbol{\Phi}_Z' + \boldsymbol{\Sigma}_Z$ and $\boldsymbol{\Gamma}_X = \boldsymbol{\Phi}_X\boldsymbol{\Gamma}_X\boldsymbol{\Phi}_X' + \boldsymbol{\Sigma}_X$ according to the Yule-Walker relationship, respectively.

Fig 1 illustrates the evolution process of $\{\mathbf{Z}_t, \mathbf{X}_t, \mathbf{Y}_t\}$, which can be viewed as a nonlinear state space model. This hierarchical model allows for flexible cross-correlations and autocorrelations of multivariate counts. It ties multiple counts together but allows for individual stochastic components through the terms $\mathbf{s}_t$ and $\boldsymbol{\epsilon}_t$. Higher-order autocorrelations of $\mathbf{Z}_t$ and $\mathbf{X}_t$ can also be accommodated in the model. Hereafter we call the proposed model as the state space multivariate Zero Inflated Poisson model (SSZIP).

$$Y_{ti}|Z_{ti}, X_{ti} = \begin{cases} 0, \text{with probability} \frac{\exp(Z_{ti})}{\exp Z_{ti}+1} \\ \text{Poisson}(\exp(X_{ti})), \text{with probability} \frac{1}{\exp Z_{ti}+1} \end{cases}$$

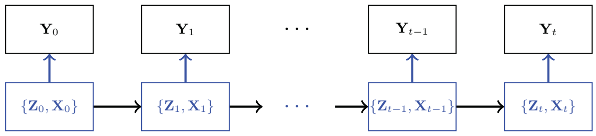

Fig. 1: The evolution of SSZIP.

III. Parameter Estimation and Inference

A. Prediction & Inference

Now we study how to make inference and predictions based on SSZIP. Here we temporarily assume that the model parameters $\boldsymbol{\Theta}$ are known, and will discuss how to estimate them in Section III-B.

Because in SSZIP, $\{\mathbf{Y}_{1:T}\}$ are observable while $\{\mathbf{Z}_{1:T}, \mathbf{X}_{1:T}\}$ are latent, most of the inference problems focus on estimating the latent process $p_{\boldsymbol{\Theta}}(\mathbf{X}_{1:T}, \mathbf{Z}_{1:T}|\mathbf{Y}_{1:T})$ given the total T observations $\{\mathbf{Y}_{1:T}\}$. More specifically, we first focus on predicting $\{\mathbf{Z}_{t+1}, \mathbf{X}_{t+1}\}$ based on the previous and current observations $\{\mathbf{Y}_{1:t}\}$, i.e., $p_{\boldsymbol{\Theta}}(\mathbf{Z}_{t+1}, \mathbf{X}_{t+1}|\mathbf{Y}_{1:t}) = p_{\boldsymbol{\Theta}}(\mathbf{Z}_{t+1}|\mathbf{Y}_{1:t}) \times p(\mathbf{X}_{t+1}|\mathbf{Y}_{1:t})$. Secondly, we study the inference of $\mathbf{X}_t$ based on the total observations $\{\mathbf{Y}_{1:T}\}$, i.e., $p_{\boldsymbol{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:T}) = p_{\boldsymbol{\Theta}}(\mathbf{Z}_t|\mathbf{Y}_{1:T}) \times p_{\boldsymbol{\Theta}}(\mathbf{X}_t|\mathbf{Y}_{1:T})$ for $t = 1, \ldots, T$. The challenge involved is because the posterior distributions $p_{\boldsymbol{\Theta}}(\mathbf{Z}_t|\mathbf{Y}_{1:t})$ and $p_{\boldsymbol{\Theta}}(\mathbf{X}_t|\mathbf{Y}_{1:t})$ has no close form for arbitrary $\boldsymbol{\Theta}$, we need to resort to numerical methods, such as numerical integration or Markov chain Monte Carlo (MCMC). In addition, we often need to update the predictions or estimations upon arrivals of new observations, e.g., to obtain $p_{\boldsymbol{\Theta}}(\mathbf{X}_{t+2}|\mathbf{Y}_{1:(t+1)})$ from $p_{\boldsymbol{\Theta}}(\mathbf{X}_{t+1}|\mathbf{Y}_{1:t})$, and $p_{\boldsymbol{\Theta}}(\mathbf{Z}_{t+2}|\mathbf{Z}_{1:(t+1)})$ from $p_{\boldsymbol{\Theta}}(\mathbf{Z}_{t+1}|\mathbf{Z}_{1:t})$. As a result, it is desirable to have computationally efficient algorithms to make inference and predictions sequentially.

Considering the performance and computational efficiency, in state space model analysis, particle filtering (PF, also called Sequential Monte Carlo) and smoothing (PS) methods are particularly useful. PF is designed to approximate $p_{\boldsymbol{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:t})$ in a sequential way with acceptably small computational complexity. Its basic idea is to compute $p_{\boldsymbol{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:t})$ by *importance sampling*, i.e., approximating $p_{\boldsymbol{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:t})$ by a set of samples, called *particles* with their associated weights. The weight assigned to each particle is proportional to its probability of being sampled from the posterior distribution. When new data are observed, new particles and their associated weights can be efficiently obtained with affordable computational burden.

In particular, for the prediction task, we assume at time t, the posterior density is approximated by weighted Dirac delta functions as

$$p_{\mathbf{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:t}) \approx \sum_{i=1}^{N_p} W_t^i \cdot \delta(\mathbf{Z}_t - \mathbf{z}_t^i)\delta(\mathbf{X}_t - \mathbf{x}_t^i). \quad (7)$$

Here in (7), δ is the Dirac delta function; N_p is the number of particles; The normalized weights W_t^i satisfy $\sum_{i=1}^{N_p} W_t^i = 1$. To obtain the prediction distribution $p_{\mathbf{\Theta}}(\mathbf{Z}_{t+1}, \mathbf{X}_{t+1}|\mathbf{Y}_{1:t})$, we can first generate samples of $\{\mathbf{Z}_{t+1}, \mathbf{X}_{t+1}\}$ from $p_{\mathbf{\Theta}}(\mathbf{Z}_{t+1}, \mathbf{X}_{t+1}|\mathbf{Y}_{1:t})$. For this purpose, each particle $\{\mathbf{z}_{t+1}^i, \mathbf{x}_{t+1}^i\}$ is propagated following the state equation in (5) with a random noise $\{\mathbf{s}_{t+1}^i, \boldsymbol{\epsilon}_{t+1}^i\}$ drawn from the state noise distribution $\mathbf{N}(\mathbf{0}, \mathbf{\Sigma}_Z)$ and $\mathbf{N}(\mathbf{0}, \mathbf{\Sigma}_X)$, i.e., $\mathbf{z}_{t+1}^i = \boldsymbol{\mu}_Z + \mathbf{\Phi}_Z(\mathbf{z}_t^i - \boldsymbol{\mu}_Z) + \mathbf{s}_{t+1}^i$ and $\mathbf{x}_{t+1}^i = \boldsymbol{\mu}_X + \mathbf{\Phi}_X(\mathbf{x}_t^i - \boldsymbol{\mu}_X) + \boldsymbol{\epsilon}_{t+1}^i$. Then we have

$$p_{\mathbf{\Theta}}(\mathbf{Z}_{t+1}, \mathbf{X}_{t+1}|\mathbf{Y}_{1:t}) \approx \sum_{i=1}^{N_p} W_t^i \cdot \delta(\mathbf{Z}_{t+1} - \mathbf{z}_{t+1}^i) \delta(\mathbf{X}_{t+1} - \mathbf{x}_{t+1}^i). \quad (8)$$

When the new observation $\mathbf{Y}_{t+1}$ comes, we can update the conditional distribution of $\{\mathbf{X}_{t+1}\}$ and approximate $p_{\mathbf{\Theta}}(\mathbf{Z}_{t+1}, \mathbf{X}_{t+1}|\mathbf{Y}_{1:t+1})$. In fact, the updated distribution takes the same form as (8) and uses the same set of particles. It only needs to update every particle's weight based on the likelihood $p_{\mathbf{\Theta}}(\mathbf{Y}_{t+1}|\mathbf{z}_{t+1}^i, \mathbf{x}_{t+1}^i)$ according to the Bayes rule, i.e.,

$$p_{\mathbf{\Theta}}(\mathbf{Z}_{t+1}, \mathbf{X}_{t+1}|\mathbf{Y}_{1:(t+1)}) \approx \sum_{i=1}^{N_p} W_{t+1}^i \cdot \delta(\mathbf{Z}_{t+1} - \mathbf{z}_{t+1}^i) \delta(\mathbf{X}_{t+1} - \mathbf{x}_{t+1}^i),$$

where $W_{t+1}^i \propto W_t^i \cdot p(\mathbf{Y}_{t+1}|\mathbf{x}_{t+1}^i)$. The convergence of the approximated distribution by PF to the true $p_{\mathbf{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:t})$ is guaranteed by the central limit theorem [15] which ensures its estimation accuracy.

One problem of PF is that the distribution of the particles' weights becomes more and more skewed as t increases. Hence, after some iterations, only very few particles have non-zero weights. This phenomenon is called *degeneracy*. We can evaluate it by the so-called effective sample size (ESS) [15], which is given by ESS $= (\sum_{i=1}^{N_p}(W_t^i)^2)^{-1}$. An intuitive solution for degeneracy is to multiply the particles with higher normalized weights, and discard the particles with lower weights. This can be done by adding a resampling step. Specifically, if ESS is smaller than a pre-specified threshold α, we resample from the set $\{(W_t^i, \{\mathbf{z}_t, \mathbf{x}_t\}^i), i = 1, \ldots, N_p\}$ with the probabilities $p(\{\hat{\mathbf{z}}_t, \hat{\mathbf{x}}_t\}^j = \{\mathbf{z}_t, \mathbf{x}_t\}^i) = W_t^i, i = 1, \ldots, N_p$ with replacement N_p times, to get a new set $\{(\frac{1}{N_p}, \{\hat{\mathbf{z}}_t, \hat{\mathbf{x}}_t\}^j), j = 1, \ldots, N_p\}$. In this way the skewness of the weights' distribution can be reduced. The detailed particle filtering procedure involving the resampling step is summarized in Algorithm 1.

Now we consider estimating the latent process given the total T observations, i.e., $p_{\mathbf{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:T}), t = 1, \ldots, T$, by PS. Its idea is to approximate $p_{\mathbf{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:T})$ with the same particles as filtering but readjust their weights by considering the information of the future observations $\{\mathbf{Y}_{t+1:T}\}$, i.e.,

$$p_{\mathbf{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:T}) \approx \sum_{i=1}^{N_p} W_{t|T}^i \delta(\mathbf{Z}_t - \mathbf{z}_t^i)\delta(\mathbf{X}_t - \mathbf{x}_t^i), \quad (9)$$

for $t = 1, \ldots, T$, where

$$W_{t|T}^i = W_t^i \sum_{j=1}^{N_p} \frac{W_{t+1|T}^j p_{\mathbf{\Theta}}(\mathbf{z}_{t+1}^j|\mathbf{z}_t^i) p_{\mathbf{\Theta}}(\mathbf{x}_{t+1}^j|\mathbf{x}_t^i)}{\sum_{l=1}^{N_p} W_t^l p_{\mathbf{\Theta}}(\mathbf{z}_{t+1}^j|\mathbf{z}_t^l) p_{\mathbf{\Theta}}(\mathbf{x}_{t+1}^j|\mathbf{x}_t^l)}, \quad (10)$$

for $t = 1, \ldots, T-1, i = 1, \ldots, N_p$, and $W_{T|T}^i = W_T^i, i = 1, \ldots, N_p$. The detailed particle smoothing procedure are summarized in Algorithm 2.

Algorithm 1 Particle filtering (PF)

At time t = 1

1: Initialization: sample $\mathbf{z}_1^i \sim p_{z0}(\mathbf{Z}_t)$ and $\mathbf{x}_1^i \sim p_{x0}(\mathbf{X}_t)$ for $i = 1, \ldots, N_p$.

2: Compute the weights $w_1^i = p(\mathbf{Y}_1|\mathbf{z}_1^i, \mathbf{x}_1^i)$ for $i = 1, \ldots, N_p$ and normalize the weights $W_1^i = \frac{w_1^i}{\sum_{i=1}^{N_p} w_1^i}, i = 1, \ldots, N_p$.

3: Calculate the filtered distribution $p(\mathbf{Z}_1, \mathbf{X}_1|\mathbf{Y}_1) \approx \sum_{i=1}^{Np} W_1^i \delta(\mathbf{Z}_1 - \mathbf{z}_1^i)\delta(\mathbf{X}_1 - \mathbf{x}_1^i)$.

At time t ≥ 2

4: Sample $\mathbf{z}_t^i \sim p_{\mathbf{\Theta}}(\mathbf{Z}_t|\mathbf{z}_{t-1}^i)$ and $\mathbf{x}_t^i \sim p_{\mathbf{\Theta}}(\mathbf{X}_t|\mathbf{x}_{t-1}^i)$ for $i = 1, \ldots, N_p$.

5: Compute the weights $w_t^i = W_{t-1}^i \cdot p(\mathbf{Y}_t|\mathbf{z}_t^i, \mathbf{x}_t^i)$ for $i = 1, \ldots, N_p$, and normalize the weights $W_t^i = \frac{w_t^i}{\sum_{i=1}^{Np} w_t^i}, i = 1, \ldots, N_p$.

6: Calculate the filtered distribution $p_{\mathbf{\Theta}}(\mathbf{Z}_t, \mathbf{X}_t|\mathbf{Y}_{1:t}) \approx \sum_{i=1}^{Np} W_t^i \delta(\mathbf{Z}_t - \mathbf{z}_t^i)\delta(\mathbf{X}_t - \mathbf{x}_t^i)$.

7: If the resample criterion satisfied, i.e., ESS $= (\sum_{i=1}^{N_p}(W_t^i)^2)^{-1} < a$, then resample with replacement N_p times from $\{\{\mathbf{z}_t, \mathbf{x}_t\}^i, i = 1 : N_p\}$ with the probabilities $p\{\hat{\mathbf{z}}, \hat{\mathbf{x}}_t\}^j = \{\mathbf{z}_t, \mathbf{x}_t\}^i = W_t^i, i = 1, \ldots, N_p$, and replace the previous set $\{(W_t^i, \{\mathbf{z}_t, \mathbf{x}_t\}^i), i = 1, \ldots, N_p\}$ by $\{(\frac{1}{N_p}, \{\hat{\mathbf{z}}_t, \hat{\mathbf{x}}_t\}^j), j = 1, \ldots, N_p\}$.

8: Terminate when $t = T$; otherwise $t = t + 1$, and go back to 4.

Algorithm 2 Particle smoothing (PS)

1: Start by setting $W_{T|T}^i = W_T^i$ for $i = 1, \ldots, N_p$.

2: For each $t = T-1, \ldots, 1$, compute the smoothed weights by

$$W_{t|T}^i = W_t^i \sum_{j=1}^{N_p} \frac{W_{t+1|T}^j p_{\mathbf{\Theta}}(\mathbf{z}_{t+1}^j|\mathbf{z}_t^i) p_{\mathbf{\Theta}}(\mathbf{x}_{t+1}^j|\mathbf{x}_t^i)}{\sum_{l=1}^{N_p} W_t^l p_{\mathbf{\Theta}}(\mathbf{z}_{t+1}^j|\mathbf{z}_t^i) p_{\mathbf{\Theta}}(\mathbf{x}_{t+1}^j|\mathbf{x}_t^l)},$$

for $i = 1, \ldots, N_p$.

3: Calculate the smoothed distribution $p_{\mathbf{\Theta}}(\mathbf{X}_t|\mathbf{Y}_{1:T}) \approx \sum_{i=1}^{N_p} W_{t|T}^i \delta(\mathbf{X}_t - \mathbf{x}_t^i)$ for $t = 1, \cdots, T$.

B. Parameter Estimation

This section considers estimation of the parameters $\mathbf{\Theta}$ of SSZIP. In the Maximum Likelihood Estimation (MLE) framework, a natural and efficient estimation method to deal with latent variables is the Expectation Maximisation (EM) algorithm. The EM algorithm is an iterative procedure to seek for $\mathbf{\Theta}^{(q)}$ in the q^{th} iteration such that the likelihood is increased from that in the $(q-1)^{th}$ iteration. Its key idea is to postulate the "missing" data $\{\mathbf{Z}_{1:T}, \mathbf{X}_{1:T}\}$ and to consider maximizing the likelihood function given the complete data $\{\mathbf{Z}_{1:T}, \mathbf{X}_{1:T}, \mathbf{Y}_{1:T}\}$. Underlying this strategy is the idea that maximizing the "complete" log-likelihood $\log p_{\mathbf{\Theta}}(\mathbf{Z}_{1:T}, \mathbf{X}_{1:T}, \mathbf{Y}_{1:T})$ is easier than maximizing the incomplete one $\log p_{\mathbf{\Theta}}(\mathbf{Y}_{1:T})$. Here due to the Markovian structure of SSZIP, the complete data log-likelihood has the form

$$\begin{aligned}\log p_{\mathbf{\Theta}}(\mathbf{Z}_{1:T}, \mathbf{X}_{1:T}, \mathbf{Y}_{1:T}) &= \log p_{z0}(\mathbf{Z}_1) + \log p_{x0}(\mathbf{X}_1) \\ &+ \sum_{t=1}^{T-1} \log p_{\mathbf{\Theta}}(\mathbf{Z}_{t+1}|\mathbf{Z}_t) + \sum_{t=1}^{T-1} \log p_{\mathbf{\Theta}}(\mathbf{X}_{t+1}|\mathbf{X}_t) \\ &+ \sum_{t=1}^{T} \log p(\mathbf{Y}_t|\mathbf{Z}_t, \mathbf{X}_t).\end{aligned} \tag{11}$$

However, because $\{\mathbf{Z}_{1:T}, \mathbf{X}_{1:T}\}$ are unavailable, the EM algorithm replaces (11) by $\mathcal{Q}(\mathbf{\Theta}, \mathbf{\Theta}^{(q)})$, which is the conditional expectation of (11) with respect to $\{\mathbf{Z}_{1:T}, \mathbf{X}_{1:T}\}$ given the observations $\{\mathbf{Y}_{1:T}\}$ using the parameters $\mathbf{\Theta}^{(q)}$ in the current iteration, i.e.,

$$\begin{aligned}\text{E step:}\quad \mathcal{Q}(\mathbf{\Theta}, \mathbf{\Theta}^{(q)}) &= \int p_{\mathbf{\Theta}^{(q)}}(\mathbf{Z}_{1:T}, \mathbf{X}_{1:T}|\mathbf{Y}_{1:T}) \\ &\cdot \log p_{\mathbf{\Theta}}(\mathbf{Z}_{1:T}, \mathbf{X}_{1:T}, \mathbf{Y}_{1:T}) \mathrm{d}\mathbf{Z}_{1:T} \mathrm{d}\mathbf{X}_{1:T}.\end{aligned} \tag{12}$$

Then we want to find the revised parameter estimates $\mathbf{\Theta}^{(q+1)}$ that maximize the function

$$\text{M step:}\quad \mathbf{\Theta}^{(q+1)} = \arg\max_{\mathbf{\Theta}} \mathcal{Q}(\mathbf{\Theta}, \mathbf{\Theta}^{(q)}). \tag{13}$$

For SSZIP, we can get $\mathcal{Q}(\mathbf{\Theta}, \mathbf{\Theta}^{(q)})$ in (12) as

$$\begin{aligned}&\mathcal{Q}(\mathbf{\Theta}, \mathbf{\Theta}^{(q)}) = \\ &\sum_{t=1}^{T-1} \iint p_{\mathbf{\Theta}^{(q)}}(\mathbf{Z}_t, \mathbf{Z}_{t+1}|\mathbf{Y}_{1:T}) \cdot \log p_{\mathbf{\Theta}}(\mathbf{Z}_{t+1}|\mathbf{Z}_t) \mathrm{d}\mathbf{Z}_t \mathrm{d}\mathbf{Z}_{t+1} \\ &+ \sum_{t=1}^{T-1} \iint p_{\mathbf{\Theta}^{(q)}}(\mathbf{X}_t, \mathbf{X}_{t+1}|\mathbf{Y}_{1:T}) \cdot \log p_{\mathbf{\Theta}}(\mathbf{X}_{t+1}|\mathbf{X}_t) \mathrm{d}\mathbf{X}_t \mathrm{d}\mathbf{X}_{t+1}.\end{aligned}$$

Unfortunately here $p_{\mathbf{\Theta}^{(q)}}(\mathbf{Z}_t, \mathbf{Z}_{t+1}|\mathbf{Y}_{1:T})$ and $p_{\mathbf{\Theta}^{(q)}}(\mathbf{X}_t, \mathbf{X}_{t+1}|\mathbf{Y}_{1:T})$ is not analytical and consequently $\mathcal{Q}(\mathbf{\Theta}, \mathbf{\Theta}^{(q)})$ is intractable. However, on the other hand, the particles used in PF & PS can be viewed as samples from the conditional distribution. As a result, we can use these particles to approximate $p_{\mathbf{\Theta}^{(q)}}(\mathbf{Z}_t, \mathbf{Z}_{t+1}|\mathbf{Y}_{1:T})$ and $p_{\mathbf{\Theta}^{(q)}}(\mathbf{X}_t, \mathbf{X}_{t+1}|\mathbf{Y}_{1:T})$, and consequently to implement the Monte Carlo EM (MCEM) algorithm for parameter estimation. In more details, from (7) and (9), we can get

$$\begin{aligned}\hat{\mathcal{Q}}(\mathbf{\Theta}, \mathbf{\Theta}^{(q)}) &\approx \sum_{t=1}^{T-1}\sum_{i=1}^{N_p}\sum_{j=1}^{N_p} \\ &W_{t,t+1|T}^{ij} \left[\log p_{\mathbf{\Theta}}(\mathbf{z}_{t+1}^j|\mathbf{z}_t^i) + \log p_{\mathbf{\Theta}}(\mathbf{x}_{t+1}^j|\mathbf{x}_t^i)\right],\end{aligned} \tag{14}$$

where

$$W_{t,t+1|T}^{ij} = W_t^i \frac{W_{t+1|T}^j p_{\mathbf{\Theta}^{(q)}}(\mathbf{x}_{t+1}^j|\mathbf{x}_t^i)}{\sum_{l=1}^{N_p} W_t^l p_{\mathbf{\Theta}^{(q)}}(\mathbf{x}_{t+1}^j|\mathbf{x}_t^l)}$$

.

In the M step, with the gradient available for (14), we get $\mathbf{\Theta}^{(q+1)} = \{\boldsymbol{\mu}_Z^{(q+1)}, \boldsymbol{\mu}_X^{(q+1)}, \mathbf{\Phi}_Z^{(q+1)}, \mathbf{\Phi}_X^{(q+1)}, \mathbf{\Sigma}_Z^{(q+1)}, \mathbf{\Sigma}_X^{(q+1)}\}$ as

$$\begin{aligned}\mathbf{\Pi}_Z^{(q+1)} &= [(\mathbf{I} - \mathbf{\Phi}_Z^{(q+1)})\boldsymbol{\mu}_Z^{(q+1)}, \mathbf{\Phi}_Z^{(q+1)}]' \\ &= (\sum_{t=1}^{T-1}\sum_{i=1}^{N_p}\sum_{j=1}^{N_p} W_{t,t+1|T}^{ij} \mathbf{z}_{t+1}^j \mathbf{u}_t^{i'})(\sum_{t=1}^{T-1}\sum_{i=1}^{N_p}\sum_{j=1}^{N_p} W_{t,t+1|T}^{ij} \mathbf{u}_t^i \mathbf{u}_t^{i'})^{-1},\end{aligned} \tag{15}$$

$$\begin{aligned}\mathbf{\Sigma}_Z^{(q+1)} = \frac{1}{T-1}\sum_{t=1}^{T-1}\sum_{i=1}^{N_p}\sum_{j=1}^{N_p} W_{t,t+1|T}^{ij} (\mathbf{x}_{t+1}^j - \mathbf{\Pi}^{(q+1)}\mathbf{u}_t^i) \\ (\mathbf{x}_{t+1}^j - \mathbf{\Pi}^{(q+1)}\mathbf{u}_t^i)'\end{aligned} \tag{16}$$

and

$$\begin{aligned}\mathbf{\Pi}_X^{(q+1)} &= [(\mathbf{I} - \mathbf{\Phi}_X^{(q+1)})\boldsymbol{\mu}_X^{(q+1)}, \mathbf{\Phi}_X^{(q+1)}]' \\ &= (\sum_{t=1}^{T-1}\sum_{i=1}^{N_p}\sum_{j=1}^{N_p} W_{t,t+1|T}^{ij} \mathbf{x}_{t+1}^j \mathbf{v}_t^{i'})(\sum_{t=1}^{T-1}\sum_{i=1}^{N_p}\sum_{j=1}^{N_p} W_{t,t+1|T}^{ij} \mathbf{v}_t^i \mathbf{v}_t^{i'})^{-1},\end{aligned} \tag{17}$$

$$\begin{aligned}\mathbf{\Sigma}_Z^{(q+1)} = \frac{1}{T-1}\sum_{t=1}^{T-1}\sum_{i=1}^{N_p}\sum_{j=1}^{N_p} W_{t,t+1|T}^{ij} (\mathbf{x}_{t+1}^j - \mathbf{\Pi}^{(q+1)}\mathbf{v}_t^i) \\ (\mathbf{x}_{t+1}^j - \mathbf{\Pi}^{(q+1)}\mathbf{v}_t^i)'\end{aligned} \tag{18}$$

where $\mathbf{u}_t^i = [1, \mathbf{z}_t^i]$ and $\mathbf{v}_t^i = [1, \mathbf{x}_t^i]$. The procedure of particle EM-based estimation is summarized in Algorithm 3.

Algorithm 3 Monte Carlo EM algorithm (MCEM)

1: Set $q = 0$ and initialize $\mathbf{\Theta}^{(q)}$ such that $\log p_{\mathbf{\Theta}^{(q)}}(\mathbf{Y}_{1:T})$ is finite.
2: Expectation (E) Step:
- Run Algorithm 1 and 2 to obtain the filtered and smoothed distributions for $t = 1, \ldots, T$.
- Calculate $\hat{\mathcal{Q}}(\mathbf{\Theta}, \mathbf{\Theta}^{(q)})$ according to (14).

3: Maximisation (M) Step:
- Compute $\mathbf{\Theta}^{(q+1)}$ according to (17) and (18) .

4: Check the non-termination condition $\hat{\mathcal{Q}}(\mathbf{\Theta}^{(q+1)}, \mathbf{\Theta}^{(q)}) - \hat{\mathcal{Q}}(\mathbf{\Theta}^{(q)}, \mathbf{\Theta}^{(q)}) \geq \epsilon$ for some $\epsilon \geq 0$. If satisfied, update $q \to q+1$, and return to 2, otherwise terminate.

IV. Numerical Studies

Here we use some numerical studies to demonstrate the proposed model and the estimation algorithm. We consider a 2-dimensional SSZIP process with series length $T = 200$. The parameters are set to be $\mathbf{\Phi}_Z = \mathbf{\Phi}_X = [0.4, 0.2; 0.2, 0.4]$, $\boldsymbol{\mu}_Z = [0, 0]$, $\boldsymbol{\mu}_X = [2, 2]$, and $\mathbf{\Sigma}_Z = \mathbf{\Sigma}_Z = 0.25\mathbf{I}$. We repeat the simulation totally 60 times. Each replicate includes data generation, estimation, and prediction, to evaluate the performance of the proposed method. For every replicate, we randomly pick $N = 12$ initial values of $\mathbf{\Theta}^{(0)}$ and estimate the parameters iteratively based on the MCEM algorithm with $N_p = 1000$ particles. Fig 2 shows the EM iteration process of a replicate. Table I lists the estimation results. We observe that both the bias and the rooted mean square error (RMSE) of the estimators are acceptably small, illustrating the satisfactory estimation accuracy and stability of the MCEM algorithm. Then we use the estimated parameters $\boldsymbol{\Theta}$ to track the latent states $\{\mathbf{Z}_{T+1:T+100}\}$ and $\{\mathbf{X}_{T+1:T+100}\}$ for the subsequent 100 observations $\{\mathbf{Y}_{T+1:T+100}\}$ according to Algorithm 1. The filtering results in one replicate are shown in Fig 3 and Fig 4. We can see that the tracked states (red crosses) based on PF almost overlap with the true states (the blue circles) with slight differences. We also do simulations for higher dimensions. Their similar results demonstrate the efficiency of the estimation procedure as well.

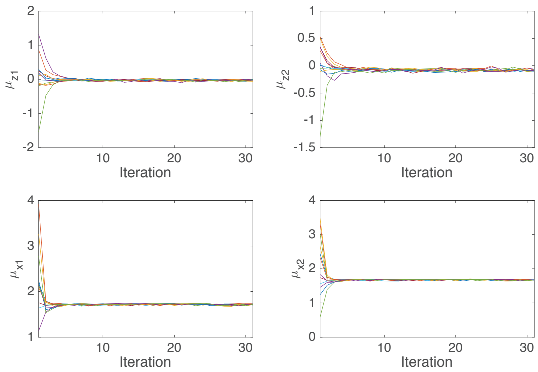

Fig. 2: MCEM iterations for the SSZIP parameters based on N = 12 replicates.

TABLE I: Parameter estimation by MCEM Algorithm based on 60 replications.

Para.	True	Bias(RMSE)	Para.	True	Bias(RMSE)
ϕ_{Z1}	0.4	-0.054 (0.082)	ϕ_{X1}	0.4	-0.006 (0.071)
ϕ_{Z2}	0.2	-0.031 (0.071)	ϕ_{X2}	0.2	0.035 (0.073)
ϕ_{Z3}	0.2	-0.039 (0.079)	ϕ_{X3}	0.2	0.013 (0.079)
ϕ_{Z4}	0.4	-0.074 (0.082)	ϕ_{X4}	0.4	0.002 (0.071)
μ_{Z1}	0	-0.036 (0.084)	μ_{X1}	2	-0.089(0.078)
μ_{Z2}	0	-0.033 (0.081)	μ_{X2}	2	-0.089 (0.074)
σ_{Z11}	0.25	-0.056 (0.061)	σ_{X11}	0.25	-0.059 (0.041)
σ_{Z12}	0	-0.021 (0.055)	σ_{X12}	0	0.026 (0.040)
σ_{Z22}	0.25	-0.061 (0.059)	σ_{X22}	0.25	0.061 (0.038)

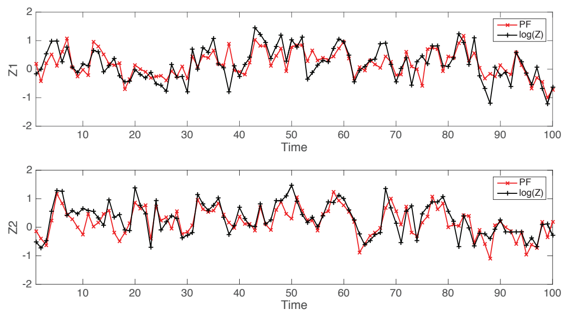

Fig. 3: The tracked states $\{\mathbf{Z}_{T+1:T+100}\}$ based on Algorithm 1 with parameters estimated by Algorithm 3.

V. CONCLUSION

This paper first proposes a state space model to describe the multivariate autocorrelated zero-inflated Poisson

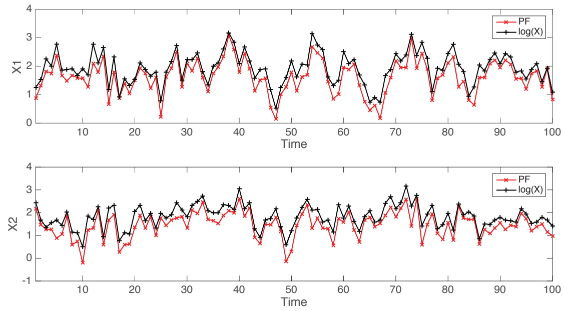

Fig. 4: The tracked states $\{\mathbf{X}_{T+1:T+100}\}$ based on Algorithm 1 with parameters estimated by Algorithm 3.

counts. This model allows for flexible zero inflations, cross-correlations and autocorrelations of multiple counts. A stable and efficient estimation procedure for this model is provided based on the MCEM algorithm together with particle filtering (PF) and particle smoothing (PS) methods. Finally, the proposed model as well as the estimation algorithm is demonstrated in the numerical studies.

REFERENCES

[1] A. C. Cohen Jr, "Estimating the parameters of a modified poisson distribution," *Journal of the American Statistical Association*, vol. 55, no. 289, pp. 139–143, 1960.

[2] N. L. Johnson and S. Kotz, *Distribution in statistics: discrete distribution*. Wiley, 1969.

[3] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992.

[4] T. Loeys, B. Moerkerke, O. De Smet, and A. Buysse, "The analysis of zero-inflated count data: Beyond zero-inflated poisson regression." *British Journal of Mathematical and Statistical Psychology*, vol. 65, no. 1, pp. 163–180, 2012.

[5] C.-S. Li, J.-C. Lu, J. Park, K. Kim, P. A. Brinkley, and J. P. Peterson, "Multivariate zero-inflated poisson models and their applications," *Technometrics*, vol. 41, no. 1, pp. 29–38, 1999.

[6] P. Holgate, "Estimation for the bivariate poisson distribution," *Biometrika*, vol. 51, no. 1-2, pp. 241–287, 1964.

[7] J.-P. Fox, "Multivariate zero-inflated modeling with latent predictors: Modeling feedback behavior," *Computational statistics & data analysis*, vol. 68, pp. 361–374, 2013.

[8] Y. Liu and G.-L. Tian, "Type i multivariate zero-inflated poisson distribution with applications," *Computational Statistics & Data Analysis*, vol. 83, pp. 200–222, 2015.

[9] C. Zhang, G. Tian, and X. Huang, "Two new bivariate zero-inflated generalized poisson distributions with a flexible correlation structure," *Statistics, Optimization & Information Computing*, vol. 3, no. 2, pp. 105–137, 2015.

[10] M. A. Jazi, G. Jones, and C.-D. Lai, "First-order integer valued ar processes with zero inflated poisson innovations," *Journal of Time Series Analysis*, vol. 33, no. 6, pp. 954–963, 2012.

[11] C. H. Weiß, "Thinning operations for modeling time series of countsła survey," *AStA Advances in Statistical Analysis*, vol. 92, no. 3, pp. 319–341, 2008.

[12] C. Li, D. Wang, and H. Zhang, "First-order mixed integer-valued autoregressive processes with zero-inflated generalized power series innovations," *Journal of the Korean Statistical Society*, vol. 44, no. 2, pp. 232–246, 2015.

[13] A. Buu, R. Li, X. Tan, and R. A. Zucker, "Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field," *Statistics in medicine*, vol. 31, no. 29, pp. 4074–4086, 2012.

[14] J. Aitchison and C. Ho, "The multivariate poisson-log normal distribution," *Biometrika*, vol. 76, no. 4, pp. 643–653, 1989.

[15] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. Springer Science & Business Media, 2008.