Received: 2 March 2021 | Accepted: 2 April 2021

DOI: 10.1002/wics.1559

ADVANCED REVIEW

WIREs COMPUTATIONAL STATISTICS WILEY

Bayesian modeling of multivariate time series of counts

Refik Soyer | **Di Zhang**

Department of Decision Sciences, The George Washington University, Washington, District of Columbia, USA

Correspondence
Refik Soyer, Department of Decision Sciences, The George Washington University, Funger Hall 415, Washington, DC 20052, USA.
Email: soyer@gwu.edu

Edited by James E. Gentle and David W. Scott, Editors-in-Chief

Abstract

In this article, we present an overview of recent advances in Bayesian modeling and analysis of multivariate time series of counts. We discuss basic modeling strategies including integer valued autoregressive processes, multivariate Poisson time series and dynamic latent factor models. In so doing, we make a connection with univariate modeling frameworks such as dynamic generalized models, Poisson state space models with gamma evolution and present Bayesian approaches that extend these frameworks to multivariate setting. During our development, recent Bayesian approaches to the analysis of integer valued autoregressive processes and multivariate Poisson models are highlighted and concepts such as "decouple/recouple" and "common random environment" are presented. The role that these concepts play in Bayesian modeling and analysis of multivariate time series are discussed. Computational issues associated with Bayesian inference and forecasting from these models are also considered.

This article is categorized under:

Statistical and Graphical Methods of Data Analysis > Bayesian Methods and Theory

Statistical Models > Time Series Models

KEYWORDS

dynamic latent factors, INAR processes, multivariate negative binomial, multivariate Poisson, non-Gaussian state space modeling

1 | INTRODUCTION AND OVERVIEW

Time-series of counts arise in many fields including business, economics, engineering, and medicine. For instance, time series under study can be the number of arrivals to a call center during every 5 min (Aktekin & Soyer, 2011), number of shopping trips of households in a week (Aktekin et al., 2018), number of mortgages defaulted from a particular pool in a given month (see Aktekin et al., 2013), time series of crash counts of different demographic groups in different geographical areas (Hu et al., 2013), number of accidents in a given time interval (Serhiyenko et al., 2014) or the number of deaths from a specific disease in a given year (Schmidt & Pereira, 2011). As noted by Soyer (2018a) "With increasing volume of Web-based data, modeling and analysis of discrete valued time series have gained more attention" in the literature. Recent advances in discrete valued time series and their applications can be found in the volume by Davis et al. (2016).

Analysis of discrete-valued series poses modeling difficulties and computational challenges. As a result, limited attention has been given to multivariate discrete valued time series in the literature. In modeling multivariate time series of counts, the framework of *observation driven* and *parameter driven* models of Cox (1981) is still applicable for temporal correlations but one also needs to describe dependence among the components of the vector time-series. As pointed out by Aktekin et al. (2020) there are alternative strategies for modeling contemporary dependence among these components. For example, one strategy is to use mixtures of distributions as discussed in Marshall and Olkin (1988) and obtain multivariate distributions. Another strategy is to describe dependence by specifying conditional distributions as in Arnold et al. (2001). These different strategies can be used to develop different models for multivariate time series of counts.

A recent review by Karlis (2016) provides an overview of the multivariate models and discuss related statistical analysis via classical methods. For example, some of the recent observation driven models for multivariate counts include multivariate integer valued autoregressive (INAR) models of Pedeli and Karlis (2011) and Pedeli and Karlis (2020). Examples of parameter driven models include Ravishanker et al. (2014) who considered a multivariate Poisson observation model with a hierarchical dynamic setup. An alternative approach was considered by Aktekin et al. (2018) who used the concept of *common random environment* to incorporate dependence among multivariate components of the count series. Both Ravishanker et al. (2014) and Aktekin et al. (2018) developed a Bayesian state-space framework to analyze their models.

In this review article, we will focus on Bayesian modeling of multivariate time series of counts. In our development we consider three main classes of multivariate time series models for counts which have gained attention in the recent Bayesian time series literature. Under each category, we start with an overview of the literature, discuss key contributions and present Bayesian work highlighting recent computational developments such as particle filtering; see for example Singpurwalla et al. (2018), and combined linear Bayes/variational Bayes methods; see for example, Berry and West (2019). In what follows, we first introduce multivariate Poisson time series models and present recent Bayesian work. This is followed by our discussion of integer autoregressive (INAR) models and the multivariate INAR (MINAR) models. We consider Bayesian analysis of univariate models and introduce some new class of MINAR processes and develop Bayesian analysis. The third class of models are the state space multivariate count models. Our coverage in this section includes random environment models as well as latent factor models of Berry and West (2019) and the *decouple/recouple* modeling strategies of West (2020). Final section includes concluding remarks.

2 | MULTIVARIATE POISSON TIME SERIES MODELS

Multivariate Poisson distribution-based time series models have gained attention in time series during the last two decades; see for example, Ravishanker et al. (2014) and Ravishanker et al. (2016). As noted by Mahamunulu (1967) the derivation of the bivariate Poisson distribution goes back to Campbell (1932) and Teicher (1954) who considered the extension to the multivariate case. Regressions of the multivariate Poisson distribution are discussed in Mahamunulu (1967). Other properties of the multivariate Poisson distribution and its generalizations can be found in the recent review by Inouye et al. (2017).

Since evaluation of the multivariate probability mass function is cumbersome and statistical inference is challenging, most of the applications of the distribution have been restricted to the bivariate case. Karlis and Ntzoufras (2003) considered a bivariate Poisson model to describe goals scored by opposing teams in soccer games. They showed that the under their set up, difference of the goals followed a Skellam distribution; see Skellam (1946). A Bayesian analysis of the bivariate Poisson model was considered in Karlis and Ntzoufras (2006). A bivariate Poisson based time series model was considered by Koopman and Lit (2015) to forecast soccer game results. The authors used a bivariate Poisson observation model and introduced temporal correlations via the Poisson rates using a state-space representation. Inference was developed using likelihood-based methods with Monte Carlo sampling.

Tsionas (1999) discussed Bayesian analysis of a special case of the multivariate Poisson distribution using a Gibbs sampler and data augmentation methods; see for example, Gelfand and Smith (1990). A Bayesian analysis of bivariate Poisson data was considered by Karlis and Tsiamyrtzis (2008) who used Markov chain Monte Carlo methods (MCMC) methods for developing inference. As noted by the authors the form of the likelihood function causes computational inefficiencies in developing inference which is more challenging for higher dimensions of the multivariate Poisson model. They proposed a recursive method for evaluating the probability distribution and a prior using mixtures of independent gamma densities. The proposed approach enabled the authors to directly sample from the posterior distributions of the parameters. In a more recent paper, Al-Wahsh and Hussein (2020) presented Bayesian analysis of the bivariate Poisson state space model to analyze daily time series of Asthma-related ER visits.

3 | BAYESIAN MODELS FOR MULTIVARIATE POISSON TIME SERIES

3.1 | Hierarchical multivariate Poisson dynamic models

Ravishanker et al. (2014) consider a vector time series counts that follow a multivariate Poisson (MVP) distribution. They have longitudinal data of multivariate counts of dimension J, that is, $\boldsymbol{Y}_{it} = (Y_{1it}, ..., Y_{Jit})$ for $i = 1, 2, ..., m$ and $t = 1, 2, ..., n$ where Y_{jit} denotes the counts for component j observed for location i at time t.

Vector time series $\boldsymbol{Y}_{it}$ is assumed to follow a J dimensional MVP distribution with parameter vector $\boldsymbol{\lambda}_{it}$ denoted as

$$\boldsymbol{Y}_{it} \,|\, \boldsymbol{\lambda}_{it} \sim \mathrm{MVP}(\boldsymbol{\lambda}_{it}). \tag{1}$$

Given $\boldsymbol{\lambda}_{it}$'s, it is assumed that the vectors $\boldsymbol{Y}_{it}$'s are independent across locations and time periods. The authors use the MVP distribution of Karlis and Meligkotsidou (2005) with the two-way covariance structure to describe dependence among the components of vector of $\boldsymbol{Y}_{it}$. This is achieved by representing $\boldsymbol{Y}_{it}$ in terms of $q = J(J+1)/2$ independent Poisson random variables X_{kit} with parameters λ_{kit}, $k = 1, 2, ..., q$, as

$$\boldsymbol{Y}_{it}^{'} = \boldsymbol{A}\boldsymbol{X}_{it} \tag{2}$$

where $\boldsymbol{X}_{it}$ is a $q \times 1$ vector and $\boldsymbol{A}$ is a $J \times q$ matrix consists of binary values. As pointed out by Ravishanker et al. (2014), matrix $\boldsymbol{A}$ in (0.2) can be decomposed as $\boldsymbol{A} = [\boldsymbol{A}_1\ \boldsymbol{A}_2]$ where $\boldsymbol{A}_1$ is a J dimensional identity matrix and $\boldsymbol{A}_2$ is a $J \times [J(J-1)]/2$ binary matrix. In the sense of Karlis and Meligkotsidou (2005) $\boldsymbol{A}_1$ can be referred to as the matrix of "main effects" and $\boldsymbol{A}_2$ can be referred to as the matrix of "two-way covariance effects."

It follows from (1) and (2) that $E[\boldsymbol{Y}_{it}|\, \boldsymbol{\lambda}_{it}] = \boldsymbol{A}\boldsymbol{\lambda}_{it}$ and $Cov[\boldsymbol{Y}_{it}|\, \boldsymbol{\lambda}_{it}] = \boldsymbol{A}\boldsymbol{\Lambda}_{it}\boldsymbol{A}^{'}$, where $\boldsymbol{\lambda}_{it}$ is a $q \times 1$ vector and $\boldsymbol{\Lambda}_{it}$ is a $q \times q$ diagonal matrix of λ_{it}'s. Note that as a result of sharing common X_{kit}'s, the above model implies a positive dependence structure among the components of $\boldsymbol{Y}_{it}$ vector. It can be shown that each component Y_{jit} follows a univariate Poisson distribution.

To illustrate the above, we consider a simple example with $m = 1$ using the Ravishanker et al. (2014) setup and assume a bivariate time series with $J = 2$. Then we write $\boldsymbol{Y}_t^{'} = \boldsymbol{A}\boldsymbol{X}_t$ as

$$(Y_{1t}\ Y_{2t}) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{pmatrix}, \tag{3}$$

where X_{kt}'s are independent Poisson random variables with parameters λ_{kt} for $k = 1, ..., 3$. Note that in (3) $\boldsymbol{A}_1$ is the two-dimensional identity matrix and $\boldsymbol{A}_2 = (1\ \ 1)^{'}$. As noted by Ravishanker et al. (2014), in the general case each column of $\boldsymbol{A}_2$ will have two 1's and $(J-2)$ 0's. For the simple case we have

$$Cov[\boldsymbol{Y}_t|\boldsymbol{\lambda}_t] = \begin{bmatrix} \lambda_{1t}+\lambda_{3t} & \lambda_{3t} \\ \lambda_{3t} & \lambda_{2t}+\lambda_{3t} \end{bmatrix},$$

where λ_{3t} implies positive dependence of Y_{1t} and Y_{2t}. The bivariate probability distribution of Y_{1t} and Y_{2t} is given by

$$P(y_{1t}, y_{2t}) = e^{-(\lambda_{1t}+\lambda_{2t}+\lambda_{3t})} \frac{\lambda_{1t}^{y_{1t}}}{y_{1t}!} \frac{\lambda_{2t}^{y_{2t}}}{y_{2t}!} \sum_{k=0}^{min(y_{1t}, y_{2t})} \binom{y_{1t}}{k} \binom{y_{2t}}{k} k! \left(\frac{\lambda_{3t}}{\lambda_{1t}\lambda_{2t}} \right)^k \tag{4}$$

where $\lambda_{kt} > 0$, $k = 1, ..., 3$ and $y_{1t}, y_{2t} = 0, 1, ...$ It can be shown that the marginals for Y_{jt}'s are Poisson with parameters $(\lambda_{jt} + \lambda_{3t})$ for $j = 1, 2$. Note that for $\lambda_{3t} = 0$ the bivariate distribution (4) reduces to the product of two independent Poisson probability mass functions.

Ravishanker et al. (2014) consider a hierarchical version of the above model $m > 1$ using $\boldsymbol{Y}_{it}$ for $i = 1, \ldots, m$. They assume that each component of the rate vector $\boldsymbol{\lambda}_{it}$ depends on exogeneous predictors with "location-time" varying coefficients via a logarithmic link function in the sense of Gamerman (1998). Temporal correlations of the model are created by introducing an "aggregate" state parameter which follows Markovian evolution in the sense of West and Harrison (1997) with Gaussian errors. Thus, the proposed structure suggests a parameter driven time-series model in the sense of Cox (1981) where component dependence of $\boldsymbol{Y}_{it}$ is achieved by the MVP model. Ravishanker et al. (2014) refer to this class of models as the *hierarchical multivariate dynamic model* (HMDM).

Given the observed vector time series $\boldsymbol{Y}_{it}$ from m locations, Bayesian analysis of the HMDM requires use of MCMC methods. The proposed approach by the authors involves using a Gibbs sampler with Metropolis-Hasting steps; see Chib and Greenberg (1995), for location-time specific parameters. The aggregate state parameter can be updated for all time periods using the Forward Filtering Backward Sampling (FFBS) of Fruhwirth-Schnatter (1994) and Carter and Kohn (1994) at each iteration of the Gibbs sampler. The authors showed the implementation of the HMDM using ecological data on gastropod counts in Puerto Rico.

3.2 | Finite mixtures of multivariate Poisson time series

One of the earlier uses of mixtures of MVPs is due to Brijs et al. (2004) who identified clusters of shoppers based on their purchasing frequency of certain type of product categories. A more detailed study of properties of MVP mixtures are given Karlis and Meligkotsidou (2007) who pointed out that the finite mixtures of MVPs have the ability to incorporate negative correlations between the components. As noted by the authors, this is an advantage of MVP mixtures over the standard MVP model which can only incorporate positive correlations.

More recently, in time series literature, using the Bayesian framework of Hu (2012), Ravishanker et al. (2016) considered mixtures of dynamic multivariate Poisson models for vector of counts from different locations. More specifically, the authors assumed a finite mixture J dimensional MVP time series for location i as

$$P(\boldsymbol{Y}_{it}|\boldsymbol{\lambda}_{it},\boldsymbol{\pi}) = \sum_{h=1}^{H} \pi_h MVP(\boldsymbol{\lambda}_{ith}) \tag{5}$$

where $\boldsymbol{\lambda}_{ith} = (\lambda_{1ith}, \lambda_{2ith}, \ldots, \lambda_{Kith})$ and $\boldsymbol{\lambda}_{it} = (\boldsymbol{\lambda}_{it1}, \boldsymbol{\lambda}_{it2}, \ldots, \boldsymbol{\lambda}_{itH})$ for locations $i = 1, \ldots, m$. Note that in (5) $MVP(\boldsymbol{\lambda}_{ith})$ is the h component of the mixture with probability π_h and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_H)$. They considered exogeneous predictors with "location-time" varying coefficients in describing λ_{kith}'s via logarithmic link functions similar to the setup given by Ravishanker et al. (2014). Markovian evolutions are assumed for the intercept terms of link functions and Bayesian analysis of the *multivariate dynamic finite mixture* (MDFM) Poisson model is developed using MCMC methods with data augmentation along the lines of Diebolt and Robert (1994).

Ravishanker et al. (2016) and Serhiyenko et al. (2018) considered a (MDFM) Poisson model for prescription counts of different drugs from different physicians. Their MVP model had $J = 3$ components with mean vector $\boldsymbol{\lambda}_{ith}$ for mixture component h for physician i. Note that at time t mean vector $\boldsymbol{\lambda}_{ith}$ has $K = 6$ elements where λ_{4ith}, λ_{5ith} and λ_{6ith} are the pairwise covariance components for the prescription counts of the three drugs. For each mixture component h, the independent Poisson parameters are defined as

$$log(\lambda_{kith}) = \begin{cases} \beta_{0kth} + \beta_{1kih} log(D_{kit}+1) + \beta_{2kih} log(Y_{ki,t-1,h}+1), & k = 1, \ldots J \\ \beta_{0kth}, & k = J+1, \ldots, K, \end{cases} \tag{6}$$

where D_{kit} is the detailing effort associated with a particular drug for physician i. We note that the means associated with the $J = 3$ drugs are dependent on D_{kit} as well as on $Y_{ki,t-1,\,h}$, the number of prescriptions written for that drug at time $t-1$ by physician i. In (6), the λ_{kith} terms for $k > J$ which are associated with the pairwise covariances do not depend on these two components. In other words, the authors consider different models for the "main effects" and the "two-way covariance effects" in the sense of Karlis and Meligkotsidou (2005).

It is also important to note that since λ_{kith}'s for $k \leq J$ are dependent on the corresponding counts $Y_{ki,t-1,\,h}$, the model by Serhiyenko et al. (2018) is an example of observation driven time series model. In fact, the implied temporal dependence by the proposed link relation is similar to the observation driven Poisson time series models considered by Davis et al. (2003).

Since the only dynamic coefficients are the intercept terms β_{0kth} in the above, model specification is completed by introducing state equations for β_{0kth} as in Serhiyenko et al. (2018) and Bayesian inference is performed using MCMC following Ravishanker et al. (2016).

As an alternative to the dynamic MVP models, Serhiyenko et al. (2018) proposed "*level correlated*" models of Serhiyenko et al. (2015) to provide a more flexible correlation structure for components of the vector of counts. The authors consider J-variate (*different drugs*) vector of counts collected on n subjects (*doctors*) over T time periods. The level correlated model (LCM) is given by

$$Y_{jit} \mid \lambda_{jit} \sim Pois(\lambda_{jit}), \tag{7}$$

for $j = 1, ..., J$, $i = 1, ..., n$, $t = 1, ..., T$ with link equation

$$log(\lambda_{jit}) = \gamma_{jt} + \beta_{ji0} + \boldsymbol{z}'_{\boldsymbol{jit}} \boldsymbol{\beta}_{ji} + \alpha_{jit}$$

where the J dimensional random effect term vector $\alpha_{it} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ induces the correlation structure for components of the count vector $\boldsymbol{Y}_{it}$. Furthermore, the state equation for each component is given by

$$\gamma_{jt} = \gamma_{j,t-1} + w_{jt}$$

where $w_{jt} \sim N(0, 1/W_j)$. This Markovian structure introduces the temporal correlations in the model.

Authors develop a Bayesian inference for the model via the integrated Laplace approximation (INLA). They point out that for large m INLA is more efficient than MCMC methods. The proposed estimation procedure seems to be quite efficient for low values of J. Although the application involves $J = 3$, it is important to note that implementation of INLA may be challenging as J increases due to evaluation of posterior modes.

Serhiyenko et al. (2018) point out that the observation equation in (7) can be generalized for any other univariate distribution of counts such as negative binomial and zero inflated Poisson models.

3.3 | Multivariate integer autoregressive processes

Development of integer valued AR (INAR) processes goes back to 1980s. Poisson AR(1) process was introduced by McKenzie (1985, 1988) and Al-Osh and Alzaid (1987).

Consider a stationary time-series Y_t represented as

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t \tag{8}$$

where "$\circ$" denotes binomial thinning operation defined as

$$\alpha \circ Y_{t-1} = \sum_{j=1}^{Y_{t-1}} B_{jt} \tag{9}$$

where B_{jt}'s are Bernoulli random variables with success probability α. Given Y_{t-1}, $\alpha \circ Y_{t-1}$ has a binomial distribution with parameters α and Y_{t-1}. In the above setup it is assumed that $\{\varepsilon_t\}$ is a sequence of iid Poisson random variables with parameter λ and B_{jt}'s are independent of ε_t's. It is important to note that thinning is performed at each time point t and is independent of thinning at other periods.

If Y_0 is assumed to be Poisson with parameter $\lambda/(1-\alpha)$ then it can be shown that Y_t's is a stationary Poisson series with parameter $\lambda/(1-\alpha)$. Also, the autocorrelation function of the process is given by $\rho_Y(k) = \alpha^k$ for $k > 0$; see for example, McKenzie (1988).

As noted by Weiß (2008), the INAR(1) can be interpreted as a special case of branching processes with immigration. Then in (8), Y_t is the population at time t which consists of two components: $\alpha \circ Y_{t-1}$, those who survive from time $(t-1)$ with probability α and ε_t, those who arrive at the beginning of time t.

Given Y_{t-1}, the conditional distribution of Y_t is obtained as a convolution of a binomial and a Poisson given by

$$p(Y_t|Y_{t-1},\lambda,\alpha) = \sum_{j=0}^{min(Y_{t-1},Y_t)} \frac{e^{-\lambda}\lambda^{Y_t-j}}{(Y_t-j)!}\binom{Y_{t-1}}{j}\alpha^j(1-\alpha)^{Y_{t-1}-j}, \tag{10}$$

with $E[Y_t|\, Y_{t-1}, \lambda, \alpha] = \alpha Y_{t-1} + \lambda$.

The k-step ahead distribution at time t, $p(Y_{t+h}\,|\,Y_t, \lambda, \alpha)$, was obtained by Silva et al. (2009) as well as the conditional mean

$$E[Y_{t+h}, Y_t, \lambda|\alpha] = \alpha^h\left(Y_t - \frac{\lambda}{(1-\alpha)}\right) + \frac{\lambda}{(1-\alpha)}. \tag{11}$$

3.4 | Bayesian analysis of univariate INAR processes

Bayesian analysis of the INAR(1) model was considered in Silva et al. (2005) using independent beta and gamma priors for α and λ, respectively. The authors used a Gibbs sampler with Metropolis since the full posterior conditionals were not known forms. Neal and Subba Rao (2007) discussed Bayesian analysis of integer ARMA processes with Poisson errors using MCMC methods which also involved use of Metropolis-Hasting steps in Gibbs sampler.

More recently, Marques et al. (2020) proposed a data augmentation algorithm which enabled them to obtain full posterior conditionals and to develop a Gibbs sampler. More specifically, by using latent variables $M_t = \alpha \circ Y_{t-1}$ where

$$M_t \,|\, Y_{t-1}, \alpha \sim Bin(Y_{t-1}, \alpha) \tag{12}$$

and the fact that $Y_t\,|\,M_t$, λ is a truncated Poisson, that is,

$$p(Y_t|M_t,\lambda) = \frac{e^{-\lambda}\lambda^{Y_t-M_t}}{(Y_t-M_t)!}\mathcal{I}(Y_t \geq M_t),$$

they obtain an augmented likelihood function for α and λ as

$$L(\lambda,\alpha;M^n,Y^n) = \prod_{t=2}^{n} p(M_t|Y_{t-1},\alpha)p(Y_t|M_t,\lambda),$$

where $M^n = (M_1, ..., M_n)$ and $Y^n = (Y_1, ..., Y_n)$.

Using independent beta and gamma priors for α and λ as

$$p(\alpha,\lambda) = Beta(a_\alpha, b_\alpha) \times Gam(a_\lambda, b_\lambda) \tag{13}$$

the full posterior conditionals can be obtained as

$$\alpha \,|\, M^n, Y^n \sim Beta\left(a_\alpha + \sum_{t=2}^{n} M_t, b_\alpha + \sum_{t=2}^{n}(Y_{t-1} - M_t)\right) \tag{14}$$

$$\lambda \mid M^n, Y^n \sim Gam\left(a_\lambda + \sum_{t=2}^{n}(Y_t - M_t), b_\lambda + (n-1)\right). \tag{15}$$

The posterior conditionals of M_t's are also obtained as

$$p(M_t|Y_t, Y_{t-1}, \alpha, \lambda) \propto \frac{1}{(Y_{t-1}-M_t)!(Y_t - M_t)!M_t!}\left(\frac{\alpha}{\lambda(1-\alpha)}\right)^{M_t} \tag{16}$$

where $M_t = 0, 1, ..., min(Y_{t-1}, Y_t)$.

3.5 | Multivariate INAR process models

Multivariate INAR processes have been considered in the literature; see for example, Latour (1997) and Pedeli and Karlis (2011, 2013). For example, a multivariate INAR(1) is discussed in Pedeli and Karlis (2013) using the "*random matricial operator*" of Latour (1997) as

$$\boldsymbol{Y}_t = \boldsymbol{A} \circ \boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{17}$$

where $\boldsymbol{Y}_t$ is a $J \times 1$ vector of time series, $\boldsymbol{\varepsilon}_t$ is the corresponding error vector and $\boldsymbol{A}$ is a $J \times J$ matrix and $\mathbf{A} \circ \mathbf{Y}_{t-1}$ is the matricial thinning operator.

As noted by Pedeli and Karlis (2020) each element α_{ij} of $\boldsymbol{A}$ in (17) defines a binomial thinning operator $\alpha_{ij} \circ Y$. For example, in the bivariate case we have

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \circ \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{18}$$

where the jth component of $\boldsymbol{Y}_t$ is given by

$$Y_{jt} = \sum_{i=1}^{2} \alpha_{ij} \circ Y_{i,t-1} + \varepsilon_{jt}$$

for $j = 1, 2$. Each component $\alpha_{ij} \circ Y_{i,t-1}$ in (18) represents a binomial thinning and as in the univariate case all thinning operations are mutually independent at time t. They are also independent of thinning at other periods and the components of error vector $\boldsymbol{\varepsilon}_t$. Thus, each component $\alpha_{ij} \circ Y_{i,t-1}$ will have a binomial distribution with parameters α_{ij} and $Y_{i,t-1}$.

Elements of error vector $\boldsymbol{\varepsilon}_t$ will have a joint distribution to reflect a dependence structure or they may be assumed to be independent. For example, Pedeli and Karlis (2020) assumed that ε_{jt}'s are independent Poisson random variables with parameters λ_j and showed that the conditional distribution of $\boldsymbol{Y}_t$ given $\boldsymbol{Y}_{t-1}$ can be obtained as the convolutions of sum of J binomials with Poissons. As shown by Pedeli and Karlis (2020), in this case, the joint distribution is the product of J generalized Poisson distributions.

A special case of the multivariate INAR(1) arises when $\boldsymbol{A}$ is diagonal in (17). In this case, since the individual series are decoupled, the dependence between Y_{jt}'s are obtained by specifying a dependence structure for elements of error vector $\boldsymbol{\varepsilon}_t$. Pedeli and Karlis (2011) considered a bivariate INAR(1) with a diagonal $\boldsymbol{A}$ matrix and assumed a bivariate Poisson model for $(\varepsilon_{1t}, \varepsilon_{2t})$ as in (4).

Estimation of the diagonal bivariate AR(1) process via likelihood-based methods are discussed in Pedeli and Karlis (2013). As noted by Karlis (2016) a Bayesian of this model was considered in the unpublished work of Sofronas (2012). Silva et al. (2005) considered replicates of INAR(1) processes and discussed the Bayesian analysis.

In what follows, we will propose an alternative modeling approach to multivariate INAR processes and develop Bayesian analysis. Our proposed approach can be considered as the multivariate generalization of the work by Marques et al. (2020).

3.6 | A Bayesian multivariate INAR process

Assume that we have J time series of counts that are conditionally independent at time t where each is described by an INAR(1) process as

$$Y_{jt} = \alpha_j \circ Y_{j,t-1} + \varepsilon_{jt} \tag{19}$$

where "$\circ$" denotes binomial thinning operation for series j as in the univariate case (8).

Given $Y_{j,t-1}$, $\alpha_j \circ Y_{j,t-1}$ has a binomial distribution with parameters α_j and $Y_{j,t-1}$. We now assume that $\{\varepsilon_{jt}\}$ is a sequence of iid Poisson random variables with parameter $\eta\lambda_j$ and ε_{jt} is independent of $\alpha_j \circ Y_{j,t-1}$. As in the univariate case, for each series thinning is performed at each time point t and is independent of thinning at other periods. Furthermore, thinning operations for each series and ε_{jt}'s are conditionally independent of other series given the respective α_j's and λ_j's and the common component η.

The *shared component* η, which follows a gamma distribution $\eta \sim Gam(a_\eta, b_\eta)$, induces dependence between the individual series. Similar to the development presented by Marques et al. (2020) we can develop Bayesian analysis for the model given data observed from J series $Y = \left(Y_1^n, ..., Y_J^n\right)$ where $Y_j^n = \left(Y_{j1}, ..., Y_{jn}\right)$.

We can introduce latent variables for each series as $M_{jt} = \alpha_j \circ Y_{j,t-1}$ where

$$M_{jt} \mid Y_{j,t-1}, \alpha_j \sim Bin\left(Y_{j,t-1}, \alpha_j\right)$$

and

$$p\left(Y_{jt} | M_{jt}, \lambda_j, \eta\right) = \frac{e^{-\eta\lambda_j}\left(\eta\lambda_j\right)^{Y_{jt}-M_{jt}}}{\left(Y_{jt}-M_{jt}\right)!}\mathcal{I}\left(Y_{jt} \geq M_{jt}\right).$$

Given the conditional independence of the series, we can write down the augmented likelihood function as

$$\prod_{j=1}^{J}\prod_{t=2}^{n} p\left(M_{jt}|Y_{j,t-1}, \alpha_j\right) p\left(Y_{jt}|M_{jt}, \lambda_j, \eta\right). \tag{20}$$

Using independent beta priors for α_j's as $\alpha_j \sim Beta\left(a_{\alpha_j}, b_{\alpha_j}\right)$ and independent gamma priors for λ_j's as $\lambda_j \sim Gam\left(a_{\lambda_j}, b_{\lambda_j}\right)$ and assuming independence of these from η, we can obtain the full posterior conditionals of all parameters.

We can show that the full conditional of η can be obtained as

$$p(\eta \mid \lambda, Y, M) \propto \left(\prod_{j=1}^{J}\prod_{t=2}^{n} e^{-\eta\lambda_j}\eta^{Y_{jt}-M_{jt}}\right) e^{-b_\eta \eta} \eta^{a_\eta - 1} \tag{21}$$

where $M = \left(M_1^n, ..., M_J^n\right)$ with $M_j^n = \left(M_{j1}, ..., M_{jn}\right)$ and $\lambda = (\lambda_1, ..., \lambda_J)$. It follows from (21) that

$$(\eta \mid \lambda, Y, M) \sim Gam\left(a_\eta + \sum_{j=1}^{J}\sum_{t=2}^{n}\left(Y_{jt} - M_{jt}\right), b_\eta + (n-1)\sum_{j=1}^{J}\lambda_j\right). \tag{22}$$

The full conditionals of α_j's and λ_j's are given by

$$\alpha_j \mid M_j^n, Y_j^n \sim Beta\left(a_{\alpha_j} + \sum_{t=2}^{n} M_{jt}, b_{\alpha_j} + \sum_{t=2}^{n} \left(Y_{j,t-1} - M_{jt}\right)\right) \tag{23}$$

and

$$\lambda_j \mid \eta, M_j^n, Y_j^n \sim Gam\left(a_{\lambda_j} + \sum_{t=2}^{n} \left(Y_{jt} - M_{jt}\right), b_{\lambda_j} + \eta(n-1)\right). \tag{24}$$

Finally, for the latent variables we can obtain

$$p\left(M_{jt} \middle| Y_{jt}, Y_{j,t-1}, \alpha_j, \lambda_j, \eta\right) \propto \frac{1}{\left(Y_{j,t-1} - M_{jt}\right)!\left(Y_{jt} - M_{jt}\right)! M_{jt}!} \left(\frac{\alpha_j}{\eta \lambda_j \left(1 - \alpha_j\right)}\right)^{M_{jt}} \tag{25}$$

where $M_{jt} = 0, 1, ..., min(Y_{j,\, t-1}, Y_{jt})$.

Note that given η for each component j of the vector time series, α_j, λ_j and M_{jt}'s are sampled independently of the other components. Similarly, forecast distributions for future values of each Y_{jt} can be obtained using standard Monte Carlo approximations based on the posterior samples from these parameters.

4 | STATE-SPACE MODELS FOR MULTIVARIATE COUNT DATA

4.1 | Earlier work

Earlier work on Bayesian modeling of time series of counts dates back to 1980s. Among these, particularly, the works of West et al. (1985) and Harvey and Fernandes (1989) contributed to the development of recent work on multivariate count models such as Aktekin et al. (2018), Berry and West (2019) and Berry et al. (2020). We first give an overview these two approaches for modeling time series of counts.

4.2 | Bayesian dynamic generalized linear models

West et al. (1985) considered dynamic version of generalized linear models from a Bayesian perspective and introduced the dynamic generalized linear model (DGLM) framework for forecasting following Harrison and Stevens (1976). Their approach includes any member of the exponential family as an observation model for time series and thus, includes binomial and Poisson time series. Linear Bayesian methods of Hartigan (1969) plays a crucial role for state variable updating in their development.

In their setup, the observation model is Poisson with mean λ_t, denoted as $Y_t \mid \lambda_t \sim Pois(\lambda_t)$. Using a logarithmic link function λ_t is related to covariate vector $\boldsymbol{F}_t$ as

$$\eta_t = log(\lambda_t) = \boldsymbol{F}_t' \boldsymbol{\theta}_t \tag{26}$$

with state equation

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t \tag{27}$$

where distribution of state error vector is partially specified via its first and second moments as $\boldsymbol{w}_t \sim [\boldsymbol{0}, \boldsymbol{W}_t]$.

Similarly, at time $t-1$ given $D^{t-1} = (Y_1, ..., Y_{t-1})$ distribution of $\boldsymbol{\theta}_{t-1}$ is specified by its first two moments as $\boldsymbol{\theta}_{t-1} \mid D^{t-1} \sim [\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}]$. Then the state equation provides us with a partially specified prior distribution for $\boldsymbol{\theta}_t$ as $\boldsymbol{\theta}_t \mid D^{t-1} \sim [\boldsymbol{a}_t, \boldsymbol{R}_t]$, where $\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1}$ and $\boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t' + \boldsymbol{W}_t$.

Using this information with the logarithmic link function we can obtain the first two prior moments of η_t given D^{t-1} as $E(\eta_t|D^{t-1}) = f_t = \boldsymbol{F}_t'\boldsymbol{a}_t$ and $V(\eta_t|D^{t-1}) = q_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t$.

We can assume a gamma prior for λ_t given D^{t-1}, or equivalently a log-gamma prior for η_t, with parameters (α_t, β_t) which can be evaluated by matching the first two moments of the log-gamma distribution with f_t and q_t. In other words, one can use the identities

$$\begin{aligned} \psi(\alpha_t) - log(\beta_t) &= f_t \\ \psi'(\alpha_t) &= q_t \end{aligned} \tag{28}$$

to solve for α_t and β_t, where $\psi(\cdot)$ is the digamma function and $\psi'(\cdot)$ is its derivative.

Once the prior parameters are specified, the posterior distribution of η_t given $D^t = (Y_t, D^{t-1})$ can be obtained as a log-gamma distribution with parameters $(\alpha_t + Y_t, \beta_t + 1)$ with posterior moments $E(\eta_t| D^t) = g_t = \psi(\alpha_t + Y_t) - log(\beta_t + 1)$ and $p_t = \psi'(\alpha_t + Y_t)$. The next step is to update the state vector $\boldsymbol{\theta}_t$ given the new observation Y_t. In WHM, this is achieved using a linear Bayesian approach which provides us with a partially specified posterior distribution. Specifically, they show that $\boldsymbol{\theta}_t \,|\, D^t \sim [\boldsymbol{m}_t, \boldsymbol{C}_t]$ where

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{R}_t\boldsymbol{F}_t(g_t - f_t)/q_t \\ \boldsymbol{C}_t &= \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{R}_t(1 - p_t/q_t)/q_t, \end{aligned} \tag{29}$$

which completes the updating from $(t-1)$ to t.

The attractive feature of the WHM approach is the availability of one-step ahead forecast (predictive) distributions at each time point as a result of using a conjugate gamma prior for the Poisson observation model. It can be easily shown that the one-step ahead forecast distribution will be obtained as a negative binomial denoted as

$$Y_t \,|\, D^{t-1}\, NBin\left(\alpha_t, \frac{\beta_t}{\beta_t + 1}\right). \tag{30}$$

4.3 | Poisson models with gamma evolution

Harvey and Fernandes (1989) also considered state space structure where the observation model is Poisson, that is, $Y_t \,|\, \lambda_t \sim Pois(\lambda_t)$. Although, their approach was not fully Bayesian, they exploited conjugacy in Bayesian updating. More specifically, given D^{t-1}, the authors assume that $\lambda_{t-1} \,|\, D^{t-1} \sim Gam(\alpha_{t-1}, \beta_{t-1})$. For specifying the prior of λ_t given D^{t-1}, to preserve the same mean but to have a higher variance, they assume a gamma distribution with parameters $\gamma\alpha_{t-1}$ and $\gamma\beta_{t-1}$ where $0 < \gamma < 1$. They noted that such a transition can be justified by the Markovian evolution

$$\lambda_t = \lambda_{t-1}\varepsilon_t/\gamma, \tag{31}$$

where $\varepsilon_t \,|\, D^{t-1} \sim Beta(\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1})$ as shown by Smith and Miller (1986) following Bather (1965). In other words, using a multiplicative transition equation with a beta distributed error term conjugacy can be preserved in the state space model.

The posterior distribution of λ_t can be obtained as a gamma distribution, that is, $\lambda_t \,|\, D^t \sim Gam(\alpha_t, \beta_t)$, where

$$\begin{aligned} \alpha_t &= \gamma\alpha_{t-1} + Y_t \\ \beta_t &= \gamma\beta_{t-1} + 1. \end{aligned} \tag{32}$$

Similar to West et al. (1985), the one step-ahead forecast distribution is given as a negative binomial. Using this fact, Harvey and Fernandes (1989) estimated γ using likelihood-based methods and also considered incorporating covariates into the model and presented maximum likelihood estimation.

Extensions of this model was considered by Aktekin and Soyer (2011) who used covariates to capture seasonal effects in call center modeling. The authors developed Bayesian analysis using MCMC methods. A similar framework later considered by Gamerman et al. (2013).

Harvey and Fernandes (1989) extended the strategy to the binomial, negative binomial, and multinomial count time series, but unlike the Poisson case the proposed transition from posterior to prior cannot be justified by calculus of probability in these cases. A multivariate extension of this framework was considered by Ord et al. (1993).

Cargnoni et al. (1997) considered Bayesian modeling of multinomial time series data using a state space framework and developed Bayesian inference using MCMC methods. Their model is an example of the class of *conditionally Gaussian dynamic models*.

4.4 | Latent factor DGLMs

Dynamic latent factor models have played an important role in Bayesian modeling of time series and spatial data. As pointed out by Lavine et al. (2020) this class of models have been considered in non-Gaussian settings including time series analysis of multivariate counts. One of the earlier uses of latent factors in state-space modeling is due to Lopes et al. (2011) who considered multivariate modeling of space–time Poisson counts. Recent work in this area such as Berry and West (2019) and Berry et al. (2020) have shown that latent factors can be used in linking multiple DGLMs of count series in implementation of "*decouple/recouple*" modeling strategy; see for example, West (2020).

Consider J Poisson count time series Y_{jt} with rate λ_{jt} denoted as $Y_{jt} \mid \lambda_{jt} \sim Poi(\lambda_{jt})$ for $j = 1, \ldots, J$. Following Lavine et al. (2020) for each series we consider a dynamic latent factor model $\mathcal{M}_j$ where λ_{jt} is related to predictor vector $\mathbf{F}_{jt}$ by a log link function as

$$\eta_{jt} = log(\lambda_{jt}) = \mathbf{F}'_{jt}\boldsymbol{\theta}_{jt}. \tag{33}$$

The state equation for $\boldsymbol{\theta}_{jt}$ is specified as

$$\boldsymbol{\theta}_{jt} = \mathbf{G}_{jt}\boldsymbol{\theta}_{j,t-1} + \mathbf{w}_{jt} \tag{34}$$

where $\mathbf{G}_{jt}$ is a known evolution matrix and the distribution of state error vector is partially specified via its first and second moments as $\mathbf{w}_{jt} \sim [\mathbf{0}, \mathbf{W}_{jt}]$. Thus, each series Y_{jt} follows a DGLM in the sense of West et al. (1985) as previously discussed.

The predictor vector $\mathbf{F}_{jt}$ consists of series specific predictors $\mathbf{f}_{jt}$ and common latent vector $\boldsymbol{\phi}_t$ as $\mathbf{F}'_{jt} = (\mathbf{f}_{jt}, \boldsymbol{\phi}'_t)$. Accordingly the state vector is defined as $\boldsymbol{\theta}'_{jt} = (\boldsymbol{\gamma}'_{jt}, \boldsymbol{\beta}'_{jt})$ allowing the possibility that each series may be affected differently by the common latent factor $\boldsymbol{\phi}_t$. This is achieved by the coefficient vector $\boldsymbol{\beta}_{jt}$ which is indexed by j in addition to being dynamic. As pointed out by West (2020) each time series Y_{jt} follows a DGLM independently of the other series, that is, given $\boldsymbol{F}_{jt}$'s, Y_{jt}'s are independent of each other. In other words, given $\boldsymbol{F}_{jt}$'s, the individual series are "decoupled."

Dependence between the individual series is achieved through the common latent vector $\boldsymbol{\phi}_t$. The latent factor $\boldsymbol{\phi}_t$ follows a model, $\mathcal{M}_0$, which allows us to "recouple" the series in the sense of West (2020). At time $t-1$ given D_0^{t-1} the external model $\mathcal{M}_0$ is specified via $p(\boldsymbol{\phi}_t|D_0^{t-1})$. Typically, model $\mathcal{M}_0$ for $\boldsymbol{\phi}_t$ is a DLM or a DGLM with its own predictors. It is assumed that we can generate samples from $p(\boldsymbol{\phi}_t|D_0^{t-1})$.

Similarly, D_j^{t-1} denotes the available count data associated with the jth component at time $t-1$ and $D_j^t = \left(D_j^{t-1}, Y_{jt}\right)$. At time $t-1$, for each series j the distribution of state vector $\boldsymbol{\theta}_{j,t-1}$ is partially described by its first two moments as $\boldsymbol{\theta}_{j,t-1} \mid D_j^{t-1} \sim \left[\boldsymbol{m}_{j,t-1}, \mathbf{C}_{j,t-1}\right]$.

Bayesian analysis of latent factor DGLM for counts consists of the forecasting and updating steps as discussed in Berry and West (2019). At time $t-1$, given samples $\boldsymbol{\phi}_t^{(s)}, s = 1, \ldots, S$, from the distribution $p(\boldsymbol{\phi}_t|D_0^{t-1})$, for each $\mathcal{M}_j$, S parallel DGLM analyses can be performed to obtain $p\left(Y_{jt}|D_j^{t-1}, \boldsymbol{\phi}_t^{(s)}\right)$. Note that $p\left(Y_{jt}|D_j^{t-1}, \boldsymbol{\phi}_t^{(s)}\right)$ is a negative binomial density as in Equation (30) of the WHM setup. As a result, for each series j,S samples $Y_{jt}^{(s)}$, $s = 1, \ldots, S$, are obtained from the predictive density $p\left(Y_{jt}|D_j^{t-1}, \boldsymbol{\phi}_t\right)$.

Note that predictive density depends on $\boldsymbol{\phi}_t$ through the link function $\eta_{jt} = \boldsymbol{F}'_{jt}\boldsymbol{\theta}_{jt}$ which is used to fit a log gamma distribution to η_{jt} with parameters $(\alpha_{jt}, \beta_{jt})$. As previously discussed, this is done by matching the first two moments of $\boldsymbol{F}'_{jt}\boldsymbol{\theta}_{jt}$ with moments of the log gamma density as shown in Equation (28). In other words, for each realization $\boldsymbol{\phi}_t^{(s)}$, a pair $\left(\alpha_{jt}^{(s)}, \beta_{jt}^{(s)}\right)$ is obtained giving us a negative binomial distribution as

$$\left(Y_{jt}|D_j^{t-1}, \boldsymbol{\phi}_t^{(s)}\right) \sim NBin\left(\alpha_{jt}^{(s)}, \frac{\beta_{jt}^{(s)}}{\beta_{jt}^{(s)}+1}\right). \tag{35}$$

Since individual series are decoupled given $\boldsymbol{\phi}_t$, the forecasting step is performed independently for each series. The k-step ahead predictive distributions can also be obtained in a similar manner as discussed in Berry and West (2019).

Once Y_{jt} is obtained at time t for each realization of $\boldsymbol{\phi}_t^{(s)}$ as implied by $\left(\alpha_{jt}^{(s)}, \beta_{jt}^{(s)}\right)$, the posterior distribution of η_{jt} is updated to a log-gamma distribution with mean $g_{jt}^{(s)}$ and $p_{jt}^{(s)}$. The linear Bayes updating is used to obtain mean vectors and variance matrices of $\boldsymbol{\theta}_{jt} \mid D_j^t, \boldsymbol{\phi}_t^{(s)}$ for $s = 1, \ldots, S$. Monte Carlo approximations are used to marginalize these over $\boldsymbol{\phi}_t^{(s)}$ to obtain the posterior moments $E\left(\boldsymbol{\theta}_{jt}|D_j^t\right)$ and $V\left(\boldsymbol{\theta}_{jt}|D_j^t\right)$. As a result the posterior moments of $\boldsymbol{\theta}_{jt}$'s are also dependent on D_0^{t-1}; see Berry and West (2019) for a more detailed discussion.

Updating of $\boldsymbol{\phi}_t$ is done externally using model $\mathcal{M}_0$. As pointed out by Berry et al. (2020), $\boldsymbol{\phi}_t$ may affect Y_{jt}'s as well as other series. For example, in their marketing applications Y_{jt}'s may represent daily sales of products $j = 1, \ldots, J$ and model $\mathcal{M}_0$ describe daily customer traffic in the store. In this context, $\boldsymbol{\phi}_t$ may represent seasonal factors affecting all series.

A flexible version of latent factor DGLMs is the *dynamic count mixture model* (DCMM) of Berry and West (2019). As noted by the authors DCMMs combine Bernoulli and Poisson DGLMs for cases where zero outcomes need to be treated separately. This is achieved by defining a binary time series $Z_{jt} = \mathcal{I}(Y_{jt} > 0)$ where $\mathcal{I}(\cdot)$ is the indicator function. The DCMM $\mathcal{M}_j$ consists of two components such that Z_{jt} is a Bernoulli denoted as $Z_{jt} \mid \pi_{jt} \sim Ber(\pi_{jt})$ and

$$Y_{jt} = \begin{cases} 0, & \text{if } Z_{jt} = 0 \\ 1 + X_{jt} & \text{if } Z_{jt} = 1, \end{cases} \tag{36}$$

where $X_{jt} \mid \lambda_{jt} \sim Pois(\lambda_{jt})$, that is, a shifted Poisson random variable for all time periods t. Under model $\mathcal{M}_j$, parameters π_{jt} and λ_{jt} are related to predictors $\boldsymbol{F}_{0jt}$ and $\boldsymbol{F}_{+jt}$ via logit and log links, respectively. As in our previous setup the predictors will consist of series specific components as well as common latent factors.

Another class of models with latent factors, where the decouple/recouple modeling strategy can be considered, are the *dynamic network flow models* (DNFMs) of Chen et al. (2018) and Chen et al. (2019). Earlier Bayesian work on network flow models considered traffic flow forecasting; see Tebaldi et al. (2002) and Anacleto et al. (2013). The recent work on DNFMs focused on large-scale "data in Internet and social network contexts."

4.5 | Random environment models

For a J component multivariate time series of counts, $\boldsymbol{Y_t} = Y_{1t}, Y_{2t}, \ldots, Y_{Jt}$, Aktekin et al. (2018) assume that these J series are exposed to the same external environment similar to common operational conditions for the components of a system as considered by Lindley and Singpurwalla (1986) in reliability analysis. For instance, in the analysis of a marketing time series, like product sales, several series are affected by the same economic swings in the market.

To account for such dependence, the authors assume that

$$\left(Y_{jt}|\lambda_j, \theta_t\right) \sim Pois\left(\lambda_j \theta_t\right), \text{for } j = 1, \ldots, J \text{ and } t = 1, \ldots, T, \tag{37}$$

where λ_j is the rate specific to the jth series and θ_t is the *common environment* modulating λ_j.

Following Harvey and Fernandes (1989), a Markovian evolution, as in (31), is assumed for θ_t, that is,

$$\theta_t = \frac{\theta_{t-1}}{\gamma}\varepsilon_t, \tag{38}$$

where, $(\varepsilon_t | D^{t-1}, \lambda_1, ..., \lambda_J) \sim Beta[\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}]$, with $\alpha_{t-1} > 0$, $0 < \gamma < 1$ and $D^{t-1} = \{D^{t-2}, Y_{1,\, t-1}, ..., Y_{J,\, t-1}\}$.

The observation model (37), is a function of both the dynamic environment θ_t as well as the static parameters, λ_js. For example, in the case where Y_{jt} represents the units of sales of brand j at time t, λ_j accounts for the effects of the brand specific rate and θ_t for the effect of the common economic environment that the all brands are exposed to at time t. When $\theta_t > 1$, the environment is said to be more favorable than usual which leads to a higher overall Poisson rate. Having the state evolution as (38) also implies a scaled beta density for θ_t where $(\theta_t | \theta_{t-1}, D^{t-1}, \boldsymbol{\lambda})$ is defined over $\left(0; \frac{\theta_{t-1}}{\gamma}\right)$ and the vector of static parameters is defined as $\boldsymbol{\lambda} = \{\lambda_1, ..., \lambda_J\}$.

In the above, we assume that for component j, given θ_t's and λ_j, Y_{jt}'s are conditionally independent over time. Furthermore, we assume that at time t, given θ_t and λ_j's, Y_{jt}'s conditionally independent of each other.

Conditional on the static parameters, it is possible to obtain an analytically tractable filtering of the states. At time 0, prior to observing any count data, we assume that $(\theta_0 | D^0) \sim Gam(\alpha_0, \beta_0)$, then it can be shown that

$$(\theta_t | D^t, \boldsymbol{\lambda}) \sim Gam(\alpha_t, \beta_t), \tag{39}$$

where

$$\alpha_t = \gamma\alpha_{t-1} + \sum_{j=1}^{J} Y_{jt}, \tag{40}$$

$$\beta_t = \gamma\beta_{t-1} + \sum_{j=1}^{J} \lambda_j. \tag{41}$$

Thus, all individual series counts as well as the all individual effects are used in updating the common random environment which plays the role of "recoupling" in the sense of Berry and West (2019).

4.5.1 | Model for multivariate counts

An important feature of the model proposed by Aktekin et al. (2018) is the availability of the marginal distribution of Y_{jt} conditional on λ_j's for $j = 1, ..., J$ which is a negative binomial given by

$$\left(Y_{jt} | \boldsymbol{\lambda}, D^{t-1}\right) \sim \mathrm{NBin}\left(\gamma\alpha_{t-1}, \frac{\lambda_j}{\gamma\beta_{t-1} + \lambda_j}\right). \tag{42}$$

As pointed out by the authors, availability of (42) in closed form is important not only for real-time forecasting purposes but also since it helps us estimating the discount factor γ.

Using the conditional independence of Y_{jt}'s, the multivariate distribution of $\boldsymbol{Y_t}$, conditional on λ_j's, can be obtained as

$$p\left(\boldsymbol{Y_t} | \boldsymbol{\lambda}, D^{t-1}\right) = \frac{\Gamma\left(\gamma\alpha_{t-1} + \sum_j Y_{jt}\right)}{\Gamma(\gamma\alpha_{t-1})\prod_j \Gamma\left(Y_{jt}+1\right)} \prod_j \left(\frac{\lambda_j}{\gamma\beta_{t-1} + \sum_j \lambda_j}\right)^{Y_{jt}} \left(\frac{\gamma\beta_{t-1}}{\gamma\beta_{t-1} + \sum_j \lambda_j}\right)^{\gamma\alpha_{t-1}}, \tag{43}$$

which is a dynamic version of a multivariate negative binomial distribution. The bivariate case of (43) with $J = 2$ is a dynamic version of the negative binomial distribution of Arbous and Kerrich (1951) who used it for modeling the number of accidents.

The conditional distributions of Y_{jt}s will also be negative binomial type where bivariate counts are positively correlated with correlation given by

$$Cor\left(Y_{it}, Y_{jt}|\lambda, D^{t-1}\right) = \sqrt{\frac{\lambda_i\lambda_j}{(\lambda_i+\gamma\beta_{t-1})(\lambda_j+\gamma\beta_{t-1})}}. \tag{44}$$

4.5.2 | Bayesian analysis of the multivariate model

Aktekin et al. (2018) discuss Bayesian inference for the multivariate model and develop details of a Gibbs sampler. They also present a particle filter as an alternative which is attractive for sequential Bayesian learning; see Soyer (2018b).

Independent gamma priors are assumed for λ_js as $\lambda_j \sim Gamma(a_j, b_j)$, for $j = 1, ..., J$. Given data D^t at time t, implementation of the Gibbs sampler requires draws from full conditionals $p(\theta_1, ..., \theta_t| \lambda_1, ..., \lambda_J, D^t)$ and $p(\lambda_1, ..., \lambda_J| \theta_1, ..., \theta_t, D^t)$. Draws from the former can be obtained using the forward filtering and backward sampling of Fruhwirth-Schnatter (1994). Since the joint density $p(\theta_1, ..., \theta_t| \lambda, D^t)$ can be factored as

$$p(\theta_t|\lambda, D^t)p\left(\theta_{t-1}|\theta_t, \lambda, D^{t-1}\right)\cdots p\left(\theta_1|\theta_2, \lambda, D^1\right),$$

with each $p(\theta_{t-1}| \theta_t, \lambda, D^{t-1})$ is a truncated gamma density with $\gamma\theta_t < \theta_{t-1}$, this can be achieved in a straightforward manner.

Drawing from $p(\lambda_1, ..., \lambda_J| \theta_1, ..., \theta_t, D^t)$ is also straightforward since λ_j's are conditionally independent given $\theta_1, ..., \theta_t$ and D^t and the marginal conditionals are gamma densities as

$$\left(\lambda_j|\theta_1, ..., \theta_t, D^t\right) \sim Gam\left(a_{jt}, b_{jt}\right), \tag{45}$$

where

$$a_{jt} = a_j + \sum_{r=1}^{t} Y_{jr}, \tag{46}$$

$$b_{jt} = b_j + \sum_{r=1}^{t} \theta_r. \tag{47}$$

As pointed out by Aktekin et al. (2018), the MCMC methods are not suitable for sequential learning and forecasting since the chains need to be restarted each time a new data vector is observed. Thus, the authors presented a particle filter (PF) as an alternative. In so doing, for sequential updating of both dynamic environment parameter and static λ_j's, they adopted the particle learning (PL) method of Carvalho et al. (2010).

The PL approach consists of *resampling*, *propagation*, *updating* and *sampling* steps. The structure of the multivariate model of Aktekin et al. (2018) provides suitable forms for the resampling weights and the propagation density required in the first two steps of the PL approach. More specifically, the weights in step one involves computing the density of *multivariate confluent hypergeometric negative binomial distribution* and the second step involves drawing samples from a univariate *scaled hypergeometric beta density* both of which can be easily done. The availability of conditional distributions for the θ_t's and λ_j's provides conditional sufficient statistics and enables authors to complete the last two steps in an efficient manner. Furthermore, since the marginal likelihood of the discount parameter γ is available as a multivariate negative binomial distribution, sequential updating of γ can also be included in the PL approach. The details of the PL algorithm and other computational issues are discussed in Aktekin et al. (2018).

4.5.3 | An extension of the random environment model

Aktekin and Soyer (2011) considered an extension of the Harvey and Fernandes (1989) to capture within day and between day correlations for arrivals to a call center. One can use a similar idea to develop an extension of the multivariate model proposed by Aktekin et al. (2018). We can assume that temporal correlations can be represented within each component by assuming time varying λ_j's. As before, we have $Y_{jt} \sim Pois(\lambda_{jt}\theta_t)$ for $j = 1, 2, ..., J$ are conditionally independent across series as well as over time.

For the random environment term θ_t we have gamma evolution model

$$\theta_t = \theta_{t-1}\varepsilon_t/\gamma$$

where $\varepsilon_t \mid \boldsymbol{\lambda}_t, D^{t-1} \sim Beta(\gamma\alpha_t, (1-\gamma)\alpha_t)$ and $\boldsymbol{\lambda}_t = (\lambda_{1t}, \lambda_{2t}, ..., \lambda_{Jt})$.

Starting with prior $(\theta_0 | D^0) \sim Gam(\alpha_0, \beta_0)$, similar to (39), we can show that $\theta_t \mid \boldsymbol{\lambda}_t, D^t \sim Gam(\alpha_t, \beta_t)$ where $\alpha_t = \gamma\alpha_{t-1} + \sum_{j=1}^{J} Y_{jt}$ and $\beta_t = \gamma\beta_{t-1} + \sum_{j=1}^{J} \lambda_{jt}$.

For the individual components, for $j = 1, 2, ..., J$ we have J independent models

$$\lambda_{jt} = \lambda_{j,t-1}\varepsilon_{jt}/\delta_j \tag{48}$$

where $\varepsilon_{jt} \mid \Theta_t, D^{j,\,t-1} \sim Beta(\delta_j a_{jt}, (1-\delta_j)a_{jt})$, and $\Theta_t = (\theta_1, ..., \theta_t)$. Note that $D^{j,\,t-1} = (Y_{j1}, ..., Y_{j,\,t-1})$ is the history of the jth component at time $(t-1)$, and a_{jt} is a known quantity at time $(t-1)$.

Assuming that $\lambda_{j0} \mid D^0 \sim Gam(a_{j0}, b_{j0})$ we can show that

$$\lambda_{jt} \mid \Theta_t, D^{jt} \sim Gam\left(a_{jt}, b_{jt}\right)$$

for $j = 1, ..., J$, where

$$a_{jt} = \delta_j a_{j,t-1} + Y_{jt} \text{ and } b_{jt} = \delta_j b_{j,t-1} + \theta_t. \tag{49}$$

Given the above results, a Gibbs sampler can be easily implemented for developing Bayesian inference for this model. Also, availability conjugate forms and the associated sufficient statistics can be exploited to develop PL methods for sequential Bayesian analysis. We note that one challenge in this model is to estimate the discount factors γ and δ_1, ..., δ_J. Using some of the recent results given by Irie et al. (2019) for updating of discount terms can be considered to learn about the discount terms.

5 | CONCLUDING REMARKS

In this review article, we have presented an overview of Bayesian modeling strategies for multivariate time series of counts. In so doing, we have highlighted recent advances in Bayesian modeling and computations for multivariate time series. Some of these recent advances have built on earlier work such as invariant conditional distributions of Bather (1965), linear Bayesian methods of Hartigan (1969), DLMs of Harrison and Stevens (1976), DGLMs of West et al. (1985) and particle filtering approach of Gordon et al. (1993). In our review we have tried to make these connections and provided the related literature.

Implementation of recent Bayesian modeling strategies for multivariate time series of counts and associated inference require use of efficient algorithms. For example, Lavine et al. (2020) discuss computational issues in dynamic for latent factor models. Implementation of PF methods and PL are discussed in Lopes and Tsay (2011) and Lopes et al. (2012).

Multivariate INAR processes is an area where Bayesian modeling strategies and inferences have not been considered. In this article, we have presented a Bayesian INAR process for multivariate time series of counts. Our proposed structure is based on a generalization of the recent work by Marques et al. (2020). Extension of the proposed approach for dynamic INAR processes is currently under investigation.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Refik Soyer: Investigation; methodology; project administration; writing-original draft. **Di Zhang:** Investigation; methodology; resources; writing-original draft.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Refik Soyer https://orcid.org/0000-0002-7854-7837

RELATED WIREs ARTICLES

Kalman Filtering and Sequential Bayesian Analysis

FURTHER READING

Meinhold, R. J., & Singpurwalla, N. D. (1983). Understanding the Kalman filter. *American Statistician*, *37*, 123–127.

Pitt, M., & Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, *94*, 590–599.

Soyer, R., & Sung, M. (2013). Bayesian dynamic probit models for analysis of longitudinal data. *Computational Statistics and Data Analysis*, *68*, 388–398.

REFERENCES

Aktekin, T., Polson, N. G., & Soyer, R. (2018). Sequential Bayesian analysis of multivariate Poisson count data. *Journal of Bayesian Analysis*, *13*, 385–409.

Aktekin, T., Polson, N. G., & Soyer, R. (2020). A family of multivariate non-Gaussian time series models. *Journal of Time Series Analysis*, *41*, 691–721.

Aktekin, T., & Soyer, R. (2011). Call center modeling: A Bayesian state space approach. *Naval Research Logistics*, *58*, 28–42.

Aktekin, T., Soyer, R., & Xu, F. (2013). Assessment of mortgage default risk via Bayesian state-space models. *Annals of Applied Statistics*, *7*, 1450–1473.

Al-Osh, M. A., & Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, *8*, 261–275.

Al-Wahsh, H., & Hussein, A. (2020). A bivariate autoregressive Poisson model and its application to asthma-related emergency room visits. *Statistics in Medicine*, *39*, 3184–3194.

Anacleto, O., Queen, C. M., & Albers, C. J. (2013). Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors. *Journal of the Royal Statistical Society (Series C: Applied Statistics)*, *62*, 251–270.

Arbous, A. G., & Kerrich, J. (1951). Accident statistics and the concept of accident proneness. *Biometrics*, *7*, 340–432.

Arnold, B. C., Castillo, E., & Sarabia, J. M. (2001). Conditionally specified distributions: An introduction. *Statistical Science*, *16*, 249–274.

Bather, J. A. (1965). Invariant conditional distributions. *Annals of Mathematical Statistics*, *36*, 829–846.

Berry, L. R., Helman, P., & West, M. (2020). Probabilistic forecasting of heterogeneous consumer transaction sales time series. *International Journal of Forecasting*, *36*, 552–569.

Berry, L. R., & West, M. (2019). Bayesian forecasting of many count-valued time series. *Journal of Business and Economic Statistics*, *38*, 872–887.

Brijs, T., Karlis, D., Swinnen, G., Vanhoof, K., Wets, G., & Manchanda, P. (2004). A multivariate Poisson mixture model for marketing applications. *Statistica Neerlandica*, *58*, 322–348.

Campbell, J. T. (1932). The Poisson correlation function. *Proceedings of Edinburgh Mathematical Society*, *4*, 18–26.

Cargnoni, C., Mueller, P., & West, M. (1997). Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association*, *92*, 640–647.

Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*, 541–553.

Carvalho, C., Johannes, M. S., Lopes, H. F., & Polson, N. G. (2010). Particle learning and smoothing. *Statistical Science*, *25*, 88–106.

Chen, X., Banks, D., & West, M. (2019). Bayesian dynamic modeling and monitoring of network flows. *Network Science*, *7*, 292–318.

Chen, X., Irie, K., Banks, D., Haslinger, R., Thomas, J., & West, M. (2018). Scalable Bayesian modeling, monitoring and analysis of dynamic network flowdata. *Journal of the American Statistical Association*, *113*, 519–533.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *American Statistician*, *49*, 327–335.

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

Davis, R., Holan, S., Lund, R., & Ravishanker, N. (2016). *Handbook of discrete-valued time series*. Chapman and Hall/CRC.

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

Diebolt, J., & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B*, *56*, 363–375.

Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, *15*, 183–202.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika*, *85*, 215–227.

Gamerman, D., Dos-Santos, T. R., & Franco, G. C. (2013). A non-Gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, *34*, 625–645.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398–409.

Gordon, N., Salmond, D., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F*, *140*, 107–113.

Harrison, P. J., & Stevens, C. F. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Series B*, *38*, 205–247.

Hartigan, J. A. (1969). Linear Bayesian methods. *Journal of the Royal Statistical Society, Series B*, *31*, 446–454.

Harvey, A. C., & Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics*, *7*, 407–417.

Hu, S. (2012). *Dynamic modeling of discrete-valued time series with applications* (Ph.D. thesis). Department of Statistics, University of Connecticut, UMI 3533942, ProQuest.

Hu, S., Ravishanker, N., Ivan, J. N., & Mooradian, J. (2013). Temporal modeling of highway crash counts for senior and non-senior drivers. *Accident Analysis and Prevention*, *50*, 1003–1013.

Inouye, D. I., Young, E., Allen, G. I., & Ravikumar, P. (2017). A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, *9*(3). https://doi.org/10.1002/wics.1398

Irie, K., Glynn, C., & Aktekin, T. (2019). *Sequential demand forecasting of bursty count data* (Under review).

Karlis, D. (2016). Models for multivariate count time series. In R. Davis, S. Holan, R. Lund, & N. Ravishanker (Eds.), *Handbook of discrete-valued time series* (pp. 407–424). CRC Press.

Karlis, D., & Meligkotsidou, L. (2005). Multivariate Poisson regression with covariance structure. *Statistics and Computing*, *15*, 255–265.

Karlis, D., & Meligkotsidou, L. (2007). Finite mixtures of multivariate Poisson regression with application. *Journal of Statistical Planning and Inference*, *137*, 1942–1960.

Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *Stat*, *52*, 381–393.

Karlis, D., & Ntzoufras, I. (2006). Bayesian analysis of the differences of count data. *Statistics in Medicine*, *25*, 1885–1905.

Karlis, D., & Tsiamyrtzis, P. (2008). Exact Bayesian modeling for bivariate Poisson data and extensions. *Statistics and Computing*, *18*, 27–40.

Koopman, S. J., & Lit, R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English premier league. *Journal of the Royal Statistical Society Series A*, *178*, 167–186.

Latour, A. (1997). The multivariate GINAR(p) process. *Advances in Applied Probability*, *29*, 228–248.

Lavine, I., Cron, A., & West, M. (2020). *Bayesian computation in dynamic latent factor models* (Technical report). Department of Statistical Science, Duke University, *ArXiv:2007.04956v1*.

Lindley, D. V., & Singpurwalla, N. D. (1986). Multivariate distributions for the life lengths of components of a system sharing a common environment. *Journal of Applied Probability*, *23*, 418–431.

Lopes, H. F., Gamerman, D., & Salazar, E. (2011). Generalized spatial dynamic factor models. *Computational Statistics and Data Analysis*, *55*, 1319–1330.

Lopes, H. F., Polson, N. G., & Carvalho, C. M. (2012). Bayesian statistics with a smile: A resampling-sampling perspective. *Brazilian Journal of Probability and Statistics*, *26*, 358–371.

Lopes, H. F., & Tsay, R. S. (2011). Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting*, *30*, 168–209.

Mahamunulu, D. M. (1967). A note on regression in the multivariate Poisson distribution. *Journal of the American Statistical Association*, *62*, 251–258.

Marques, P. C. F., Graziadei, H., & Lopes, H. F. (2020). Bayesian generalizations of the integer-valued autoregressive model. *Journal of Applied Statistics*, 1–21. https://doi.org/10.1080/02664763.2020.1812544

Marshall, A. W., & Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, *83*, 834–841.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, *21*, 645–650.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, *20*, 822–835.

Neal, P., & Subba Rao, T. (2007). MCMC for integer-valued ARMA processes. *Journal of Time Series Analysis*, *28*, 92–110.

Ord, K., Fernandes, C., & Harvey, A. C. (1993). Time series models for multivariate series of count data. *Developments in Time Series Analysis: In Honour of Maurice B. Priestley, T. Subba Rao* (pp. 295–309).

Pedeli, X., & Karlis, D. (2011). A bivariate INAR(1) model with application. *Statistical Modelling*, *11*, 325–349.

Pedeli, X., & Karlis, D. (2013). On estimation of the bivariate Poisson INAR process. *Communications in Statistics—Simulation and Computation*, *42*, 514–533.

Pedeli, X., & Karlis, D. (2020). An integer-valued time series model for multivariate surveillance. *Statistics in Medicine*, *39*, 940–954.

Ravishanker, N., Serhiyenko, V., & Willig, M. R. (2014). Hierarchical dynamic models for multivariate times series of counts. *Statistics and Its Interface*, *7*, 559–570.

Ravishanker, N., Venkatesan, R., & Hu, S. (2016). Dynamic models for time series of counts with a marketing application. In R. A. Davis, S. H. Holan, R. Lund, & N. Ravishanker (Eds.), *Handbook of discrete-valued time series* (pp. 423–445). CRC Press.

Schmidt, A. M., & Pereira, J. B. M. (2011). Modelling time series of counts in epidemiology. *International Statistical Review*, *79*, 48–69.

Serhiyenko, V., Ivan, J. H., Ravishanker, N., & Islam, S. (2014). Dynamic compositional modeling of pedestrian crash counts on urban roads in Connecticut. *Accident Analysis & Prevention*, *64*, 78–85.

Serhiyenko, V., Ravishanker, N., & Venkatesan, R. (2015). Approximate Bayesian estimation for multivariate count time series models. In P. K. Choudhary, et al. (Eds.), *Ordered data analysis, modeling and health research methods* (pp. 155–167). Springer International.

Serhiyenko, V., Ravishanker, N., & Venkatesan, R. (2018). Multi-stage multivariate modeling of temporal patterns in prescription counts for competing drugs in a therapeutic category. *Applied Stochastic Models in Business and Industry*, *34*, 61–78.

Silva, I., Silva, M. E., Pereira, I., & Silva, N. (2005). Replicated INAR(1) process. *Methodology and Computing in Applied Probability*, *7*, 517–542.

Silva, N., Pereira, I., & Silva, M. E. (2009). Forecasting in INAR(1) model. *Revstat Statistical Journal*, *7*, 119–134.

Singpurwalla, N. D., Polson, N. G., & Soyer, R. (2018). From least squares to signal processing and particle filtering. *Technometrics*, *60*, 146–160.

Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society Series A*, *109*, 296.

Smith, R., & Miller, J. E. (1986). A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society, Series B*, *48*, 79–88.

Sofronas, G. (2012). *Bayesian estimation for bivariate integer autoregressive model* (Unpublished master thesis). Department of Statistics, Athens University of Economics, Athens, Greece.

Soyer, R. (2018a). Discussion of multi-stage multivariate modeling of temporal patterns in prescription counts for computing drugs in a therapeutic category. *Applied Stochastic Models for Business & Industry*, *34*, 79–80.

Soyer, R. (2018b). Kalman filtering and sequential Bayesian analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *10*(5). https://doi.org/10.1002/wics.1438

Tebaldi, C., West, M., & Karr, A. F. (2002). Statistical analyses of freeway traffic flows. *Journal of Forecasting*, *21*, 39–68.

Teicher, H. (1954). On the multivariate Poisson distribution. *Skandinavisk Aktuarietidskrift*, *37*, 1–9.

Tsionas, E. G. (1999). Bayesian analysis of the multivariate Poisson distribution. *Communications in Statistics—Theory and Methods*, *28*, 431–451.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts—A survey. *Advances in Statistical Analysis*, *92*, 319–341.

West, M. (2020). Bayesian forecasting of multivariate time series: Scalability, structure uncertainty and decisions. *Annals of the Institute of Statistical Mathematics*, *72*, 131.

West, M., & Harrison, P. J. (1997). *Bayesian forecasting and dynamic models* (2nd ed.). Springer-Verlag.

West, M., Harrison, P. J., & Migon, H. S. (1985). Dynamic generalized linear models, and Bayesian forecasting. *Journal of the American Statistical Association*, *80*, 73–97.

How to cite this article: Soyer, R., & Zhang, D. (2021). Bayesian modeling of multivariate time series of counts. *Wiley Interdisciplinary Reviews: Computational Statistics*, e1559. https://doi.org/10.1002/wics.1559