Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Count Time Series: A Methodological Review

Richard A. Davis, Konstantinos Fokianos, Scott H. Holan, Harry Joe, James Livsey, Robert Lund, Vladas Pipiras & Nalini Ravishanker

To cite this article: Richard A. Davis, Konstantinos Fokianos, Scott H. Holan, Harry Joe, James Livsey, Robert Lund, Vladas Pipiras & Nalini Ravishanker (2021): Count Time Series: A Methodological Review, Journal of the American Statistical Association, DOI: 10.1080/01621459.2021.1904957

To link to this article: https://doi.org/10.1080/01621459.2021.1904957

Published online: 04 May 2021.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 00, NO. 0, 1–15: Review
https://doi.org/10.1080/01621459.2021.1904957

Taylor & Francis
Taylor & Francis Group

Count Time Series: A Methodological Review

Richard A. Davis[a], Konstantinos Fokianos[b], Scott H. Holan[c,d], Harry Joe[e], James Livsey[d], Robert Lund[f], Vladas Pipiras[g], and Nalini Ravishanker[h]

[a]Department of Statistics, Columbia University, New York, NY; [b]Department of Mathematics and Statistics, University of Cyprus, Nicosia, Cyprus; [c]Department of Statistics, University of Missouri, Columbia, MO; [d]U.S. Census Bureau, Washington, DC; [e]Department of Statistics, University of British Columbia, Vancouver, Canada; [f]Department of Statistics, The University of California—Santa Cruz, Santa Cruz, CA; [g]Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC; [h]Department of Statistics, University of Connecticut, Storrs, CT

ABSTRACT

A growing interest in non-Gaussian time series, particularly in series comprised of nonnegative integers (counts), is taking place in today's statistics literature. Count series naturally arise in fields, such as agriculture, economics, epidemiology, finance, geology, meteorology, and sports. Unlike stationary Gaussian series where autoregressive moving-averages are the primary modeling vehicle, no single class of models dominates the count landscape. As such, the literature has evolved somewhat ad-hocly, with different model classes being developed to tackle specific situations. This article is an attempt to summarize the current state of count time series modeling. The article first reviews models having prescribed marginal distributions, including some recent developments. This is followed by a discussion of state-space approaches. Multivariate extensions of the methods are then studied and Bayesian approaches to the problem are considered. The intent is to inform researchers and practitioners about the various types of count time series models arising in the modern literature. While estimation issues are not pursued in detail, reference to this literature is made.

ARTICLE HISTORY
Received April 2020
Accepted March 2021

KEYWORDS
Autocorrelation; Count Negative binomial distribution; Poisson distribution; State-space models; Time series

1. Introduction

Classical time series techniques, in both the frequency and time domains, are typically second-order based. That is, modeling does not go beyond the first (mean) and second (covariance) moments. As a result, Gaussian models, which are completely characterized by their first two moments, became popular. Eventually, researchers sought more, realizing that Gaussian models often poorly described count and other discrete-valued series. Count time series modeling began in earnest in the late 1970s, with work on the topic rapidly accelerating thereafter. This article overviews the types of count time series models used in the historical and modern literature. While our focus is on modeling techniques and methods, references to estimation techniques are presented.

Count time series arise in numerous applied scientific areas. Examples include the daily number of patients admitted in a hospital, the number of transactions of a given stock observed every minute, or the monthly number of car accidents in a region. These data, occasionally observed with some covariates, often share some common characteristics. Foremost, they are frequently dependent as they are observed sequentially in time. As counts, they are integer valued. Count series are often overdispersed (i.e., their variance is greater than their mean) and their autocorrelations are often nonnegative. Finally, there are often more zero counts (zero-inflation) than can be explained by the classical marginal count distributions (e.g., Poisson, binomial, negative binomial); see Alqawba, Diawara, and Rao Chaganty (2019) and Sellers, Peng, and Arab (2020) and the references therein for more on zero-inflated count models. Good models and analyses should account for these features and provide the user with estimation, fitting, model assessment, prediction, and uncertainty quantification.

Originally, count series were often described via generalized linear models. Generalized linear models can accommodate continuous or discrete observations and extend linear regression for Gaussian data to exponential family distributed data. The unification of discrete and continuous paradigms began in Nelder and Wedderburn (1972) and was further developed in McCullagh and Nelder (1989) for independent series. Early work for correlated count series can be found in Cox (1981); here, models were classified into two types: observation and parameter driven. This terminology will be explained in Section 4. Subsequent generalized linear modeling contributions include Wong (1986), Holden (1987), Zeger and Qaqish (1988), Zeger (1988), Albert et al. (1994), Li (1994), Davis, Wang, and Dunsmuir (1999), Davis, Dunsmuir, and Wang (2000), Davis, Dunsmuir, and Streett (2003), and Fokianos and Kedem (2004). Early textbooks on the topic are Fahrmeir and Tutz (2001) and Kedem and Fokianos (2002).

As the field developed, various approaches for modeling correlated count series emerged. Unlike Gaussian series, where autoregressive moving-average (ARMA) schemes dominated, many types of models, including integer ARMA, discrete ARMA, generalized ARMA, etc. were proposed over the years. All approaches had drawbacks; consequently, the field developed without a unifying theory.

CONTACT Robert Lund *rolund@ucsc.edu* Department of Statistics, The University of California - Santa Cruz, Santa Cruz, CA 95064.

The rest of this article, which is an attempt to review, summarize, and relate these models, proceeds as follows. Section 2 begins with models based on thinning operators, one of the initial modeling approaches. The theory of thinning operators is linked to strictly stationary series with a prespecified count marginal distribution, but models based on thinning operators can be extended to include covariates. Section 3 tours some recent developments in this area. Section 4 explores state-space approaches to the problem, including Markov chain and hidden Markov models. Multivariate extensions are subsequently considered in Section 5. Section 6, our last technical section, takes a Bayesian tour of count modeling. A section on future directions closes the article.

2. Count Time Series Models Based on Thinning Operators

2.1. Overview

In early count series research, the goal was often to construct strictly stationary series supported on the nonnegative integers with a variety of autocorrelation functions (ACF), and with marginal distributions such as Poisson, negative binomial, and generalized Poisson. A few references are McKenzie (1986, 1988), Alzaid and Al-Osh (1993), Al-Osh and Aly (1992), and Joe (1996). Recent reviews include Weiß (2008), Scotto, Weiß, and Gouveia (2015), and Weiß (2018).

One direction of current research lies with integer autoregressive models, denoted by INAR(p), where p is the autoregressive order. An INAR(p) or generalized INAR(p) series is constructed by summing thinning operations applied to the p past series values with an innovation variate. The distribution of the innovation is chosen from a given parametric family such as Poisson or negative binomial. The resulting marginal distribution of the process may not have a simple form. References include Du and Li (1991), Latour (1998), Gauthier and Latour (1994), and Joe (2016).

The basic idea is perhaps the best illustrated in the first-order INAR(1) setting. Here, a difference equation governs $\{X_t\}$

$$X_t = \alpha * X_{t-1} + \epsilon_t, \tag{1}$$

where $\{\epsilon_t\}$ is an IID count-valued random sequence. The operator $*$ mimics autoregressive multiplication, keeping the process count-valued. Specifically, $*$ operates via $\alpha * W = \sum_{i=1}^{W} B_i$ for any count variable W, where $\{B_i\}_{i=1}^{\infty}$ are IID Bernoulli trials, independent of W, with success probability $\alpha \in [0, 1]$. A series with Poisson(λ) marginal distributions, for example, can be achieved by taking $\{\epsilon_t\}$ to be IID Poisson with mean $(1 - \alpha)\lambda$.

Under minor irreducibility assumptions, $\{X_t\}$ in Equation (1) is a Markov chain with a unique stationary distribution. We are interested in its stationary version made by either starting the series in its stationary distribution at time zero, or by viewing the process as having started infinitely far in the past and having reached statistical equilibrium.

Empirical marginal distributions and the sample ACF of $\{X_t\}$ often suggest particular correlation structures and marginal distributions. The sample dispersion is another useful diagnostic: if ϵ_t is overdispersed (larger than unity), then under mild conditions, the marginal distribution of X_t is also overdispersed. If ϵ_t is underdispersed, then the marginal distribution could be either underdispersed or overdispersed. A Poisson $\{X_t\}$, of course, has unit dispersion.

Non-stationary extensions of these models, with a mean or other parameter being a function of covariates, exist, as do other types of thinning operators. Extensions depend on the chosen thinning operator, some being more unwieldly than others. See Section 8.4.4 of Joe (1997) for covariate extensions with thinning operators created so that marginal distributions lie in a convolution-closed family. Thinning operators have been extended to (i) operate on all integers (including negative ones), and to (ii) build multivariate count models (see Section 5).

For models based on thinning operators, estimation is done via maximum likelihood when possible; otherwise, conditional least squares, composite likelihood, or other methods merit consideration. See Graziadei, Lopes, and Marques (2020) for Bayesian approaches and Pedeli and Karlis (2013a) for composite likelihood methods.

2.2. General Thinning Operators

This subsection discusses different thinning operators and their properties, which are an active current research areas. Unless otherwise noted, the thinning operators act on the set of nonnegative integers, have a parameter $\alpha \in [0, 1]$, and map a count valued random variable X into another count valued random variable X_α such that $\mathbb{E}[X_\alpha] = \alpha\mathbb{E}[X]$ when X has a finite mean. Some thinning operators satisfy $X_\alpha \leq X$ in the sense of some stochastic ordering; examples include binomial, beta-binomial, and quasi-binomial thinning operators. For thinning operators with compounding operations, it is possible that $X_\alpha > X$; however, it is more likely that $X_\alpha \leq X$. As $\alpha \downarrow 0$, the thinning operator maps the input to zero; as $\alpha \uparrow 1$, the thinning operator does not change the input.

Let X be a nonnegative random variable with cumulative distribution function F_X such that for all $0 < \alpha < 1$, there exists an innovation random variable $\epsilon(\alpha)$, independent of X, with

$$X \overset{\mathcal{D}}{=} \alpha * X + \epsilon(\alpha).$$

Then X is called discrete self-decomposable (DSD) and the stationary count model $X_t = \alpha * X_{t-1} + \epsilon_t(\alpha)$ has marginal distribution F_X for all $0 < \alpha < 1$. It is known that DSD random variables are infinitely divisible, and hence compound Poisson, and that their dispersion is at least unity.

However, if $\alpha \in (0, 1)$ and the cumulative distribution of ϵ_t, say F_ϵ, is fixed, the recursion (dependence on α is suppressed in some notations below)

$$X_t = \alpha * X_{t-1} + \epsilon_t, \tag{2}$$

still defines a legitimate count time series model. The pair (α, F_ϵ) determines the stationary distribution F_X. If F_ϵ is underdispersed, it is possible for F_X to be underdispersed.

Binomial thinning is a special case of expectation thinning. Expectation thinning is based on the compounding of a nonnegative integer random variable $K(\alpha)$ with $\alpha \in [0, 1]$; via,

$$K(\alpha) \circledast \ell = \sum_{i=1}^{\ell} K_i(\alpha), \tag{3}$$

where $\{K_i(\alpha)\}_{i=1}^{\infty}$ are IID replicates of $K(\alpha)$. It is assumed that $\mathbb{E}[K(\alpha)] = \alpha$, $K(0) \equiv 0$ and $K(1) \equiv 1$, so that count models can be constructed similar to Equation (2).

To construct an integer analog of a Gaussian AR(p), thinning operators based on compounding are similar. A generalized integer-autoregression of order p, denoted by GINAR(p), obeys

$$X_t = \sum_{j=1}^{p} K(\alpha_j) \circledast X_{t-j} + \epsilon_t = \sum_{j=1}^{p} \sum_{i=1}^{X_{t-j}} K_{j,i,t}(\alpha_j) + \epsilon_t, \qquad (4)$$

where $\alpha_j \in [0,1]$ for $j = 1, \ldots, p$ and the $K_{j,i,t}(\alpha_j)$ are independent over i, j, and t, and ϵ_t is the innovation at time t. This can be viewed as a branching process with immigration, where a particle at time t has independent branching at times $t+1, \ldots, t+p$. Numerical maximum likelihood estimation for GINAR(p) series with covariates is considered in Joe (2019).

The most common INAR(p) form in the literature uses binomial thinning (Du and Li 1991). Because the survival/continuation interpretation for binomial thinning does not extend to the second and higher lag orders, it is better to consider Equation (4) with more general thinning operators. Families of expectation thinning operators with the closure property $K(\alpha_1) \circledast K(\alpha_2) = K(\alpha_1\alpha_2)$, for $\alpha_1, \alpha_2 \in [0,1]$, have desirable properties; see Zhu and Joe (2010).

In the stationary case, the ACF often takes the classical Gaussian AR(p) form, but does not have full range. Specifically, negative serial correlations cannot be produced with these models. GINAR(1) models have a geometric ACF with positive correlations. If Markov models of order p with negative serial correlations are desired, then approaches based on copulas applied to $p+1$ consecutive observations are available (Joe 2016); one can also use the methods in Section 3.

3. New Developments

This section narrates some recent developments in stationary count time series modeling. The goal here is again to build stationary count series having the cumulative marginal distribution F_X. Here, difference equation tactics are eschewed in attempts to produce very general count structures; in particular, the series below can have long memory and/or negative correlations. Long memory here means that $\sum_{h=0}^{\infty} |\gamma_X(h)| = \infty$, where $\gamma_X(h) = \mathrm{cov}(X_t, X_{t+h})$.

3.1. Construction from Binary Series

It is known that any discrete distribution can be generated from independent coin flips (even fair coin flips). Mimicking this idea, Blight (1989) and Cui and Lund (2009) attempt to construct series with the prescribed marginal distribution by combining correlated binary (Bernoulli) processes. Jia, Lund, and Livsey (2020b) is the most recent reference on the technique.

Let $\{B_t\}_{t=1}^{\infty}$ be a binary random series taking values in $\{0,1\}$ with lag h covariance $\gamma_B(h) = \mathrm{cov}(B_t, B_{t+h})$. The count series $\{X_t\}$ constructed will inherit properties of $\gamma_B(\cdot)$. Before constructing $\{X_t\}$, we present two methods of generating stationary correlated Bernoulli sequences.

First, a renewal-based $\{B_t\}$ is based on a "lifetime" $L \in \{1, 2, \ldots\}$ and "initial delay" $L_0 \in \{0, 1, \ldots\}$. Here, a random walk $\{S_n\}_{n=0}^{\infty}$ is defined via

$$S_n = L_0 + L_1 + \cdots + L_n, \qquad n \geq 0,$$

and the binary sequence is defined by $B_t = 1$ if $S_n = t$ for some $n \geq 0$. This is the usual renewal point process. In this setup, $L_1, L_2, \ldots$ are IID, each having the same distribution as L. If $L_0 = 0$, the process is called nondelayed; if the distribution of L_0 differs from that of L, the setup is said to have delay L_0.

To make $\{B_t\}$ stationary, L_0 must have the distribution $\mathbb{P}[L_0 = k] = \mathbb{P}(L > k)/\mathbb{E}[L]$ for $k \in \{0, 1, \ldots\}$. In this case, $\{B_t\}$ is stationary with $\gamma_B(h) = \mathbb{E}[L]^{-1}(u_h - \mathbb{E}[L]^{-1})$, where u_h denotes the probability of obtaining a renewal (point) at time h in the nondelayed process. Typical regularity conditions assume a finite $\mathbb{E}[L]$ and a support set for L that is aperiodic. Under this, $u_h \to 1/\mathbb{E}[L]$ as $h \to \infty$ and $\gamma_B(h) \to 0$ as $h \to \infty$. Lund, Holan, and Livsey (2016) showed that $\{B_t\}$ has long memory (absolutely nonsummable covariances) if and only if $\mathbb{E}[L^2] = \infty$. Specifically, when $\mathbb{E}[L^2] = \infty$, but $\mathbb{E}[L] < \infty$, $\{B_t\}$ will have long memory. Observe that $\gamma_B(h) < 0$ whenever $u_h < 1/\mathbb{E}[L]$. As negative and long-memory series can be produced, covariance structures are more flexible than those for the thinned series of the last section.

Second, a Gaussian clipped $\{B_t\}$ is built from a latent Gaussian process $\{Z_t\}$ with zero mean, unit variance, and autocovariance/ACF $\rho_Z(h) = \mathrm{Corr}(Z_t, Z_{t+h})$. Tong (1990b) is a good reference for general Gaussian sequences. The Bernoulli sequence is constructed by placing (clipping) $\{Z_t\}$ into zero-one categories: $B_t = 1_{(Z_t > 0)}$. Then $\mathbb{E}[B_t] = 1/2$ and classic quadrant integration results show that $\gamma_B(h) = \sin^{-1}(\rho_Z(h))/(2\pi)$. This process will have long memory whenever $\{Z_t\}$ has long memory and will have a negative covariance at lag h whenever $\gamma_Z(h) < 0$. The clipping threshold need not be set to zero and the paradigm easily extends to multivariate settings. Again, the obtained covariance structures are very flexible. For references, Kedem (1980) discusses general properties of clipped Gaussian processes and Livsey et al. (2018) constructs clipped count models.

Returning to constructing the count series $\{X_t\}$, we need IID copies of $\{B_t\}$ — call these $\{B_{1,t}\}, \{B_{2,t}\}, \ldots$. These binary processes are combined in a way to produce the desired marginal distribution. For example, if one wants binomial marginal distributions with M trials and success probability p, set

$$X_t = \sum_{j=1}^{M_t} B_{j,t}; \qquad (5)$$

here, $M_t \equiv M$ and L is chosen to have mean $1/p$ for renewal $\{B_t\}$ (or Z_t is clipped at a threshold rendering $\mathbb{P}(B_t = 1) = p$).

If a Poisson(λ) marginal distribution is desired, take $\{M_t\}$ to be IID Poisson with mean λ/p (a Poisson sum of independent zero-one draws is again Poisson distributed). Note that M_t is IID here; specifically, the construction is not circular. Should a negative binomial marginal distribution with positive integer trials parameter r and success probability p is desired, one can look at the first index at which r successes are obtained in the

time t Bernoulli trials:

$$X_t = \inf\left\{k \geq r : \sum_{i=1}^{k} B_{i,t} = r\right\} - r$$

(r is subtracted to make X_t lie in $\{0, 1, \ldots\}$).

While how to best combine the $B_{i,t}$'s to produce a desired marginal distribution with the widest range of autocorrelations is an open question, in general, models with flexible ACF are easily achieved. Blight (1989), Cui and Lund (2009), Fralix, Livsey, and Lund (2012), Lund and Livsey (2016), Lund, Holan, and Livsey (2016), and Livsey et al. (2018) explore issues further. Estimation is done through Gaussian quasi-likelihood with process means and autocovariances. The key property is that when $\{M_t\}$ is IID, the lag-h autocovariance of $\{X_t\}$ in Equation (5) is proportional to $u_h - 1/\mathbb{E}[L]$ in the renewal case, or $\sin^{-1}(\rho_Z(h))$ in the Gaussian case that is clipped at zero. This permits negative covariances and/or long memory features to be achieved. See Jia, Lund, and Livsey (2020b) for the latest.

3.2. Latent Gaussian Transformations

Stationary count series with the marginal distribution F_X can also be constructed through the probability integral transform and a stationary Gaussian series. This model is tantamount to using a Gaussian copula for all finite-dimensional distributions.

The Gaussian copula/transformation produces very flexible count series as one can have any count marginal distribution. The construct uses a latent zero mean, unit variance, stationary Gaussian sequence $\{Z_t\}$ with $\rho_Z(h) = \text{Corr}(Z_t, Z_{t+h})$ and transforms this via

$$X_t = F_X^{-1}(\Phi(Z_t)), \tag{6}$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution and

$$F_X^{-1}(u) = \inf\{t : F_X(t) \geq u\}, \quad u \in (0, 1), \tag{7}$$

is the generalized inverse (quantile function) of F_X. See Smith and Khaled (2012), Masarotto and Varin (2012), and Han and De Oliveira (2016) for copula techniques in spatial statistics and other settings. This model is called the normal to anything (NORTA) approach in operations research (Cario and Nelson 1997; Chen 2001) and a translational model in mechanical engineering (Grigoriu 2007). When $\{Z_t\}$ is an autoregression, Biller (2009) calls this approach the autoregressive-to-anything model.

Biller (2009) mentions other non-Gaussian copulas that can generate stationary count series; when $\{Z_t\}$ is AR(1), many copula forms have been explored. For counts, parametric copula families for Markov transition probabilities are used in (Joe 1997, chap. 7) and (Joe 2014, sec. 7.5). Escarela, Mena, and Castillo-Morales (2006) developed a copula-based Markov chain model for binary longitudinal data.

The probability integral transformation theorem shows that $\Phi(Z_t)$ has a uniform(0,1) distribution for each t; a second application of the result shows that X_t has marginal distribution F_X for each t. Temporal dependence in $\{Z_t\}$ induces temporal dependence in $\{X_t\}$. To quantify this, let $\gamma_X(h) = \text{cov}(X_t, X_{t+h})$. The autocovariances of $\{X_t\}$ and $\{Z_t\}$ can be related using Hermite expansions (see Pipiras and Taqqu 2017, chap. 5). Specifically, let $G(x) = F_X^{-1}(\Phi(x))$ and expand $G(\cdot)$ in the Hermite polynomial basis $\{H_k(\cdot)\}_{k=0}^{\infty}$ defined by

$$H_k(z) = (-1)^k e^{z^2/2} \frac{d^k}{dz^k}\left(e^{-z^2/2}\right)$$

($H_0(z) \equiv 1$, $H_1(z) = z$, and $H_2(z) = z^2 - 1$; higher order polynomials can be obtained from the recursion $H_k(z) = zH_{k-1}(z) - H'_{k-1}(z)$). The expansion has form $G(x) = \sum_{k=0}^{\infty} g_k H_k(x)$ and the Hermite coefficients are

$$g_k = \frac{1}{k!}\int_{-\infty}^{\infty} G(z)H_k(z)\frac{e^{-z^2/2}dz}{\sqrt{2\pi}} = \frac{1}{k!}\mathbb{E}[G(Z_0)H_k(Z_0)],$$

which Jia et al. (2020a) showed can be expressed as

$$g_k = \frac{1}{k!\sqrt{2\pi}} \sum_{n=0}^{\infty} e^{-\Phi^{-1}(Q_n)^2/2} H_{k-1}(\Phi^{-1}(Q_n))$$

whenever F_X has a finite $(1 + \delta)$th moment for some $\delta > 0$. Here, $Q_n = F_X(n)$.

The autocovariances of $\{X_t\}$, perhaps the primary quantity of interest, can be related to the autocorrelations of $\{Z_t\}$ via

$$\gamma_X(h) = \sum_{k=1}^{\infty} k! g_k^2 \gamma_Z(h)^k =: C(\rho_Z(h)),$$

where the power series has form $C(u) = \sum_{k=1}^{\infty} k! g_k^2 u^k$. Autocorrelations have the form

$$\rho_X(h) =: \text{corr}(X_t, X_{t+h}) = \sum_{k=1}^{\infty} \frac{k! g_k^2}{\gamma_X(0)} \gamma_Z(h)^k =: L(\rho_Z(h)),$$

where $L(u) = \sum_{k=1}^{\infty} \ell_k u^k$, with $\ell_k = k! g_k^2/\gamma_X(0)$. $L(\cdot)$ is called a link function and maps $[-1, 1]$ into (but not necessarily onto) $[-1, 1]$. Observe that $L(1) = \sum_{k=1}^{\infty} \ell_k = 1$; however, $L(-1)$ is not necessarily -1 in general. As such, $L(\cdot)$ "starts" at $(-1, L(-1))$, passes through $(0, 0)$, and connects to $(1, 1)$. Jia et al. (2020a) argued that $|\rho_X(h)| \leq |\rho_Z(h)|$, that a positive $\rho_Z(h)$ leads to a positive $\rho_X(h)$, and that a negative $\rho_Z(h)$ leads to a negative $\rho_X(h)$. This follows by establishing that $L(u)$ is monotone increasing in u and crosses zero at $u = 0$.

The autocovariance memory properties of $\{X_t\}$ follow from those of $\{Z_t\}$. In particular, if $\{Z_t\}$ is short memory, then so is $\{X_t\}$. If $\{Z_t\}$ is long memory, then so is $\{X_t\}$ provided the Hermite rank of $G(\cdot)$, which is defined as $r = \min\{k \geq 1 : g_k \neq 0\}$, is no more than unity. If $\{Z_t\}$ is q-dependent, then $\{X_t\}$ is also q-dependent.

Copula models have very flexible autocorrelations. Indeed, these methods produce the most flexible bivariate correlation, $\text{Corr}(X_{t_1}, X_{t_2})$, possible when X_{t_1} and X_{t_2} have the same marginal distribution F_X (Whitt 1976). Since a general count marginal distribution can also be achieved, the model class appears quite general. Estimation for this model class, including how to deal with covariates, is studied in Jia et al. (2020a).

4. Generalized State-Space Models

Generalized state-space modeling is one of the most common and flexible frameworks for time series analysis. For example, ARMA and ARIMA models can be cast in a linear state-space model for which Kalman filtering, smoothing, and prediction algorithms can be implemented for calculating Gaussian likelihoods, one-step predictors, mean-squared prediction errors, and smoothed values of the states. Moreover, straightforward modifications of these models can handle missing values, structural breaks, interventions, and seasonality. For count time series, the linear state-space model is not directly applicable since the response variable assumes discrete values; a more general form of the standard generalized state model is required.

For our purposes, only an informal definition of a generalized state-state space model is given; this will inject some flavor without burying the reader in notation. A state-space model consists of two specifications: one for the observed time series X_t at time t given a state process α_t, and a second that describes the evolution of $\{\alpha_t\}$ in t. While Bayesian methods will be considered more fully in Section 6, one often uses a known probability mass function (PMF) for the observation specification. For example, a Poisson model assumes

$$X_t|\alpha_t \sim \text{Poisson}(e^{\alpha_t}), \tag{8}$$

where Poisson(λ) denotes a Poisson PMF with mean λ. This is a natural extension of the Poisson regression model that has been classically used to analyze counts. While the Poisson specification is the most common for counts, it does not always provide a good fit. Other marginal PMFs include the negative binomial and generalized Poisson among others; more generally, one could assume that the conditional PMF comes from the one-parameter exponential family

$$\mathbb{P}(X_t = x_t|\alpha_t) = \exp\{\varphi(x_t) + \alpha_t x_t - A(\alpha_t)\}, \quad x_t = 0, 1, 2, \ldots, \tag{9}$$

where $\varphi(\cdot)$ is a function, and $A(\alpha)$ is a normalizing constant that ensures the PMF in (9) sums to unity: $A(\alpha) = \log(\sum_{j=0}^{\infty} \exp\{\varphi(j) + \alpha j\})$. Usually, $\varphi(x)$ is a known function of $x \in \{0, 1, \ldots, \}$, such as $-\log(x!)$ in the Poisson case. However, in Zhang (2018), $\varphi(\cdot)$ is viewed as a parameter to be estimated. From Equation (9), we have

$$B(\alpha_t) =: A'(\alpha_t) = \mathbb{E}[X_t|\alpha_t], \quad B'(\alpha_t) = \text{Var}(X_t|\alpha_t).$$

For the state-process $\{\alpha_t\}$, there are typically two approaches for modeling, commonly referred to as parameter- and observation-driven. In the parameter-driven case, it is assumed that $\{\alpha_t\}$ follows its own stochastic mechanism, which is often Markov. For the Poisson model in Equation (8), a popular model takes $\{\alpha_t\}$ as a Gaussian AR(1) process satisfying

$$\alpha_t = \phi_0 + \phi_1\alpha_{t-1} + \epsilon_t, \quad \{\epsilon_t\} \sim \text{IID } N(0, \sigma^2).$$

The parameters obey $\phi_0, \phi_1 \in (0, 1)$, and $\sigma^2 > 0$. These models can be estimated via frequentist or Bayesian approaches (see Section 6 for details).

In the more general case of the one-parameter exponential family in Equation (9), one typically uses a link function and specifies a Gaussian time series model for $\{\alpha_t\}$ that is related to the conditional mean via

$$\lambda_t =: \mathbb{E}[X_t|\alpha_t] = B(\alpha_t) \quad \text{or} \quad \alpha_t = B^{-1}(\lambda_t).$$

Since $B(\alpha)$ is strictly increasing in α ($B'(\alpha) > 0$), there is a one-to-one connection between α_t and λ_t. This fairly general approach extends the GLM family.

One can incorporate p-dimensional covariates $\mathbf{c}_t$ at time t by setting

$$\alpha_t = \mathbf{c}_t^T\boldsymbol{\beta} + Z_t, \tag{10}$$

where $\boldsymbol{\beta}$ is a p-dimensional regression parameter vector and $\{Z_t\}$ is say a stationary Gaussian process — this is a generalized linear mixed model. If $Z_t \equiv 0$, the model reduces to the standard GLM. In the Poisson case, the unconditional mean is $\mathbb{E}[X_t] = \mathbb{E}[\exp\{\mathbf{c}_t^T\boldsymbol{\beta} + Z_t\}] = \exp\{\mathbf{c}_t^T\boldsymbol{\beta}\}\mathbb{E}[e^{Z_t}] = \exp\{\mathbf{c_t}^T\boldsymbol{\beta} + \kappa\}$, where $\kappa = \text{var}(Z_t)/2$. In other words, the covariates are interpretable through the unconditional mean in exactly the same fashion as a traditional GLM and $\{Z_t\}$ accounts for overdispersion. The straightforward inclusion of interpretable covariates is but one advantage of a parameter-driven model. In fact, one can directly use GLM estimates of $\boldsymbol{\beta}$, which turn out to be consistent and asymptotically normal. However, the limiting covariance matrix has to account for $\{Z_t\}$. Details can be found in Davis, Dunsmuir, and Wang (2000) and Davis and Wu (2009). See also Zeger (1988) for one of the first parameter-driven count series models. The parameter-driven model can be viewed as a hierarchical Bayesian model (see Section 6), where the observation equation is the "data model" and the stochastic mechanism for the state process is the "process model." However, in a fully Bayes' formulation, prior distributions would be placed on both the data and process model parameters.

Model likelihood computation based on data $X_1, \ldots, X_n$ for parameter-driven models is problematic since it requires an n-fold integral over the AR(1) Gaussian probability density function for $(\alpha_1, \ldots, \alpha_n)^T$. Simulation-based MCMC and importance sampling methods, as well as approximations to the likelihood and pairwise composite likelihoods, exist for carrying out inference; however, these procedures can be challenging and are not uniformly reliable.

In contrast, observation-driven models offer an attractive alternative to parameter-driven models, especially on estimation fronts. For the observation-driven models considered here (Brockwell and Davis 2016, sec. 9.6 gives a more formal treatment), α_t is assumed to be an explicit (or implicit) function of the past observations $X_s, s < t$. A popular choice for the one-parameter exponential family (8) is one that mimics the dynamics of a GARCH process for the conditional mean $\lambda_t = \mathbb{E}[X_t|\alpha_t] = B(\alpha_t)$. This is defined through the recursion

$$\lambda_t = d + a_1\lambda_{t-1} + b_1X_{t-1}, \tag{11}$$

where d, a_1, and b_1 are all nonnegative. While this recursion involves λ_{t-1} and X_{t-1}, one can recurse backwards $t-1$ iterations to see that λ_t is indeed a function of $X_s, s = 0, \ldots, t-1$, and λ_0:

$$\lambda_t = d\sum_{i=0}^{t-1} a_1^i + a_1^t\lambda_0 + b_1\sum_{i=0}^{t-1} a_1^iX_{t-1-i}. \tag{12}$$

Recursing to the infinite past yields

$$\lambda_t = \frac{d}{1-a_1} + b_1\sum_{i=0}^{\infty} a_1^iX_{t-1-i}, \tag{13}$$

provided $|a_1| < 1$ and the series is well defined. The process $\{\lambda_t\}$ in Equation (11) is Markov; if there is a stationary solution, then using $\mathbb{E}[X_t] = \mathbb{E}[\lambda_t]$ in (11) gives $\mathbb{E}[\lambda_t] = d/(1-a_1-b_1)$ when $a_1 + b_1 < 1$. In fact, $a_1 + b_1 < 1$ is a necessary and sufficient conditions for the existence of a strictly stationary solution to Equation (11) with a finite mean. One of the inherit difficulties with observation-driven models lies with establishing stability properties (such as stationarity and mixing) of the state process. While $\{\lambda_t\}$ in (13) describes a simple relation between λ_t and its predecessor λ_{t-1}, it is not straightforward to establish the existence of a stationary distribution. Standard Markov theory results, such as ϕ-irreducibility, do not apply in this situation. This is due in large part to the state-space of the conditional distribution of $\lambda_t|\lambda_{t-1}$ depending on λ_{t-1}. Many strategies exist for overcoming these limitations; some successful efforts lie in Fokianos, Rahbek, and Tjøstheim (2009) and Neumann (2011) for the Poisson case, and in Davis and Liu (2016) for the general one-parameter exponential family case. In the latter, the theory of iterated random functions is used as developed in Diaconis and Freedman (1999) and Wu and Shao (2004). For the special case of Equation (11), if $a_1 + b_1 < 1$, then there exists a unique strictly stationary solution to the recursion and the resulting process $\{X_t\}$ is β-mixing at a geometric rate. This result is important for inference purposes, from which one can show that the maximum likelihood estimators are asymptotically normal.

4.1. Observation-Driven Models

This section expands on observation-driven models. Computations can be carried out by a number of existing software packages, including the recently released R package `tscount` (Liboschik, Fokianos, and Fried 2017) that can fit the linear and log-linear models below. These models can accommodate both positive and negative correlations with suitable model parametrizations. These issues were addressed in Zeger and Qaqish (1988) and Davis, Wang, and Dunsmuir (1999).

For count series, simple generalizations of autoregressive models are the linear and log-linear autoregressive models. The linear autoregressive Poisson model of order p obeys

$$X_t|\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \lambda_t = d + \sum_{j=1}^{p} b_j X_{t-j}, \tag{14}$$

where $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ is the process history, and d and $\{b_j\}_{j=1}^p$ are nonnegative (this guarantees a nonnegative Poisson mean). This, and all subsequent models in this section, do not imply that the marginal distribution of X_t is Poisson; indeed, the marginal distribution of X_t can be far from Poisson. However, the Poisson conditional assumptions permit easy derivation of a conditional likelihood. Quasi-likelihood methods can also be used for inference (Fokianos, Rahbek, and Tjøstheim 2009; Christou and Fokianos 2014; Ahmad and Franq 2016); in fact, Equation (14) is a generalized linear model. Elaborating, following the standard terminology in McCullagh and Nelder (1989), the random component is the Poisson distribution, the systematic component is $d + \sum_{j=1}^{p} b_j X_{t-j}$, and the link function is the identity map. It can be shown that $\mathbb{E}[X_t] = d/(1 - \sum_{j=1}^{p} b_j)$ and that the autocovariance function of $\{X_t\}$ has an identical form to that of a classical AR(p) model (Fokianos 2012).

An analogous log-linear model for count series is defined as

$$\nu_t = d + \sum_{j=1}^{p} b_j \log(X_{t-j} + 1), \tag{15}$$

where $\nu_t = \log(\lambda_t)$ (Zeger and Qaqish 1988). Here, the parameters d and $\{b_j\}_{j=1}^p$ can be positive or negative, but must obey the stationarity conditions in Sim (2016). Note that the lagged past observations of the response drive the autoregressive equation for ν_t via $\log(X_{t-j} + 1)$. This is a one-to-one transformation of X_{t-j}. In this setup, λ_t and X_t are put on the same scale. Covariates can be easily accommodated in Equation (15) via an additional additive term.

The models in Equations (14) and (15) have short memory autocovariances akin to ordinary AR(p) or ARCH(p) (autoregressive conditional heteroscedastic) models; see Engle (1982). Numerous recent studies extend the paradigm to settings where memory decays more slowly. Toward this, several authors specify a generalized ARCH (GARCH) model, where Equation (14), in the spirit of Bollerslev (1986), is replaced by

$$X_t|\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \lambda_t = d + \sum_{i=1}^{p} a_i \lambda_{t-i} + \sum_{j=1}^{q} b_j X_{t-j}, \tag{16}$$

where $a_i > 0$ for $i = 1, 2, \ldots, p$ and $b_j > 0$ for $j = 1, 2, \ldots, q$; see Rydberg and Shephard (2000), Streett (2000), Heinen (2003), Ferland, Latour, and Oraichi (2006), and Fokianos, Rahbek, and Tjøstheim (2009) among others. The model in (16) is parsimonious, but has autocovariances that decay geometrically rapidly to zero with increasing lag. When a count series' autocorrelation decays more slowly to zero, the model in Equation (14) can approximate this structure with a large p. Further remarks about Equation (16) now follow.

For the Poisson distribution, $\mathbb{E}[X_t|\mathcal{F}_{t-1}] = \text{Var}(X_t|\mathcal{F}_{t-1}) = \lambda_t$. Because of this, Equation (16) is often called an integer GARCH (INGARCH) model, since its structure mimics that of the customary GARCH model.

Similarly, one can generalize the log-linear model in Equation (15) to

$$\nu_t = d + \sum_{i=1}^{p} a_i \nu_{t-i} + \sum_{j=1}^{q} b_j \log(X_{t-j} + 1), \tag{17}$$

and its properties are analogous to Equation (16) (Fokianos and Tjøstheim 2011). No positivity constraints are required for model coefficients and covariates can be easily included. Other nonlinear models for count time series analysis have been studied in Fokianos and Tjøstheim (2012), Christou and Fokianos (2014), and Wang et al. (2014). Series properties are established in Neumann (2011), Doukhan, Fokianos, and Tjøstheim (2012), and Douc, Doukhan, and Moulines (2013). Maximum and quasi-maximum likelihood estimation for the models in Equations (16) and (17) are discussed in Fokianos, Rahbek, and Tjøstheim (2009), Christou and Fokianos (2014), Ahmad and Franq (2016), and Fokianos (2015) (among others). Extensions to general one-parameter exponential families and to nonlinear dynamics for $\{\lambda_t\}$ are studied in Davis and Liu (2016).

Threshold models allow governing parameters to change with previous observations; see Tong (1990a) for a general reference. These nonlinear models have been used for a variety of applications in biology, meteorology, and macro-economics. Assuming a Poisson PMF for X_t given λ_t, the basic threshold INGARCH model, often called the self-excited INGARCH(1,1) (SETINGARCH) model, posits that

$$\lambda_t = \begin{cases} d_1 + a_1\lambda_{t-1} + b_1 X_{t-1}, \text{ if } X_{t-1} \le r \\ d_2 + a_2\lambda_{t-1} + b_2 X_{t-1}, \text{ if } X_{t-1} > r \end{cases}, \tag{18}$$

where $d_1, d_2 > 0$, $a_1, a_2, b_1, b_2 \ge 0$, and r is the threshold parameter. The process obeys two different recursions, the one in use depending on whether or not X_{t-1} exceeds r. Unlike the INGARCH family, these models permit negative correlations at lag one. The theoretical development of this model, including maximum likelihood estimation, is developed in Wang et al. (2014). Like the corresponding Gaussian threshold AR(1) model, it is possible to have a stationary solution if a_1 and b_1 lie outside the "stationary regime" where $a_1 + b_1 < 1$. However, one must have $a_2 + b_2 < 1$ when X_{t-1} is large to probabilistically prevent the process from diverging to infinity.

The generalized linear ARMA (GLARMA) family is one of the most flexible and easily fit count models that balances parameter and observation-driven models. Assuming a one-parameter exponential family for the observation process, a zero-mean unit-variance martingale difference sequence (MGD) can be constructed from the conditional expectation $\lambda_t = \mathbb{E}[X_t|\alpha_t] = B(\alpha_t)$, that is,

$$e_t = \frac{X_t - B(\alpha_t)}{\sqrt{B'(\alpha_t)}}. \tag{19}$$

Since a martingale difference sequence is uncorrelated, it follows that $\{e_t\}$ is white noise with zero mean and a unit variance. The GLARMA(p, q) model for the data is then based on an ARMA(p, q) recursion driven by $\{e_t\}$. Specifically, let $\{W_t\}$ be the causal-invertible ARMA process obeying

$$W_t = \sum_{i=1}^{p} \phi_i \alpha_{t-i} + e_t + \sum_{j=1}^{q} \theta_j e_{t-j}.$$

Then α_t is defined as the best linear predictor of W_t given the infinite past $\{W_s, s < t\}$ (or equivalently the infinite past $\{X_s, s < t\}$), which can be shown to satisfy the ARMA-like recursions (Davis, Wang, and Dunsmuir 1999; Dunsmuir 2016)

$$\begin{aligned} \alpha_t &= \phi_1(\alpha_{t-1} + e_{t-1}) + \cdots + \phi_p(\alpha_{t-p} + e_{t-p}) \\ &\quad + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} \\ &= \sum_{i=1}^{p} \phi_i \alpha_{t-i} + \sum_{i=1}^{\tilde{q}} \tilde{\theta}_i e_{t-i}, \end{aligned} \tag{20}$$

where $\tilde{q} = \max(p, q)$ and $\tilde{\theta}_i = \phi_i + \theta_i$. Here, the conventions $\phi_i = 0$ for $i > p$ and $\theta_i = 0$ for $i > q$ are made. This is a parameter-driven model since α_t depends on previous observations. In practice, Equation (20) is initialized with $e_t = \alpha_t = 0$ or $t \le 0$. Model fitting can be carried out using the R package `glarma` (Dunsmuir et al. 2015). An advantage of this model is that covariates can be incorporated in a meaningful way. For example, one can replace α_t by $\tilde{\alpha}_t = \mathbf{c}_t^T\boldsymbol{\beta} + \alpha_t$, where $\mathbf{c}_t$ is a covariate vector at time t and $\boldsymbol{\beta}$ is an unknown regression vector. In this setting, $e_t = (X_t - B(\tilde{\alpha}_t))/\sqrt{B'(\tilde{\alpha}_t)}$. Since the mean and variance of e_t do not depend on the covariates, $\mathbb{E}[\tilde{\alpha}_t] = \mathbf{c}_t^T\boldsymbol{\beta}$ as desired.

Unfortunately, establishing strict stationarity and mixing for $\{\alpha_t\}$ has only been done in special cases (Davis, Dunsmuir, and Streett 2003, 2005). A general theory has not yet been developed, although simulation results show that standard limit theory for likelihood estimates appears to hold.

There is a close connection between GLARMA and score-driven models (sometimes referred to as generalized AR score (GAS) and dynamic conditional score (DCS) models) that are well described in Harvey (2013) and Creal, Koopman, and Lucas (2013) (see also *http://www.gasmodel.com* for further references). The score function for the one-parameter exponential family in Equation (9) is

$$\frac{\partial \log(\mathbb{P}(X_t|\alpha_t))}{\partial \alpha_t} = X_t - B(\alpha_t) = X_t - \mathbb{E}[X_t|\alpha_t].$$

Since the score has mean zero, it is a MGD sequence coinciding with the numerator in Equation (19). One can normalize this MGD by any function of α_t without altering the MGD property. The GAS model is then defined as the AR process driven by this rescaled MGD sequence. The dynamics of $\{\alpha_t\}$ obey

$$\alpha_t = d + a\alpha_{t-1} + b\left(\frac{X_{t-1} - B(\alpha_{t-1})}{S(\alpha_{t-1})}\right), \tag{21}$$

where $|a| < 1$ and $S(\cdot)$ is a specifically chosen scale function. While choosing $S(\alpha) = \sqrt{B'(\alpha)}$ ensures that the MGD sequence has a unit variance, other choices can be advantageous (Creal, Koopman, and Lucas 2013). Aside from an intercept term, this expression for α_t coincides with the GLARMA model in Equation (20), with $p = q = 1$ and $e_t = b(X_t - B(\alpha_t))/S(\alpha_t)$.

4.2. Markov and Hidden Markov Models

Markov and hidden Markov models (HMMs) and their analysis techniques often arise in the study of count series. Modeling a count series $\{X_t\}$ taking S distinct values as a general Markov chain is often unfruitful because the one-step-ahead transition matrix contains $S(S-1)$ free parameters, which is not parsimonious when S is large. This concern lessens when S is small; in fact, only two parameters are needed to model a binary-valued Markov series.

Many of the previous count series examined, which are generally parsimonious models, are in fact Markov chains, perhaps after a suitable enlargement of the state space. This is the case, for example, with INAR series and their convolution-closed extensions, and GLAR series. Here, Markov techniques have been used to establish stationarity, stability, and other series properties. See Fokianos et al. (2019) for examples. When a count series is known to be Markov, likelihood inference becomes tractable and other inference procedures can be conducted (Joe 1997, 2016; Zucchini and MacDonald 2016).

Some count series are also HMMs. By this, we mean that there is a background latent stationary Markov chain $\{Z_t\}$ evolving in time and that the count draw X_t at time t depends on Z_t, but not on past values of the latent chain. This is the case with the

latent Gaussian models in Section 3.2 when the Gaussian series $\{Z_t\}$ is AR(p) (and is thus pth-order Markov). The HMM structure results from X_t being determined (albeit deterministically in this model) from Z_t. Connections to HMMs can be exploited, for example, for inference tasks. Indeed, estimation for HMMs can be carried out via a well-studied framework involving particle filtering, smoothing, and other methods (Douc, Moulines, and Stoffer 2014). For HMM techniques with a latent Gaussian $\{Z_t\}$, see Jia et al. (2020a).

Other authors use the (H)MM framework to drive their discourse. For example, (Joe 1997, sec. 8.1.2), develops a binary Markov model taking values in $\{0, 1\}$ that has Markov transition probabilities driven by two natural parameters. Another popular HM count model assumes that the marginal distribution depends on parameter(s), with the parameter(s) evolving in a Markov fashion. For example, the marginal distribution of the series could be a mixture of two Poisson distributions, with the transitions between the two Poisson parameters governed by a Markov chain. The temporal dependence in this construction is parameter-driven, akin to the GLARMA construction in Equation (17) with $q = 0$. For more details, including estimation and numerous data examples, see, for example, Zucchini and MacDonald (2009).

5. Multivariate Models

Interest in multivariate count series is driven by a range of applied problems, including medicine (Paul, Held, and Toschke 2008), finance (Pedeli and Karlis 2013b), marketing (Ravishanker, Venkatesan, and Hu 2016), and the environment (Livsey et al. 2018). The review article Karlis (2016) provided further detail. In this section, several modeling approaches in the literature are discussed.

5.1. Multivariate INAR models

The initial multivariate approaches involved INAR models as in Franke and Rao (1995) and Latour (1997), which were recently discussed in Pedeli and Karlis (2013a,2013b), and Scotto et al. (2014). We now elaborate on the simplest case, which involves binomial thinning; however, as in Section 2.2, other thinning operators are possible. For a $d \times d$ matrix $\mathbf{A} = (\alpha_{i,j})$ of binomial thinning parameters and a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)^T$, define

$$\mathbf{A} * \mathbf{X} = \begin{pmatrix} \sum_{j=1}^{d} \alpha_{1,j} \circ X_j \\ \vdots \\ \sum_{j=1}^{d} \alpha_{d,j} \circ X_j \end{pmatrix},$$

where each thinning is performed independently of all others. The multivariate INAR(1) process $\mathbf{X}_t = (X_{1,t}, \ldots, X_{d,t})^T$ is defined as the unique (in mean square) stationary solution to

$$\mathbf{X}_t = \mathbf{A} * \mathbf{X}_{t-1} + \mathbf{E}_t, \tag{22}$$

where $\{\mathbf{E}_t\}$ is a sequence of IID count-valued d-variate random vectors. Generalizations to multivariate INAR(p) models for orders $p > 1$ simply add additional "thinned $\mathbf{X}_{t-\ell}$ lagged variables" for $\ell > 1$ in Equation (22). Estimation for multivariate INAR models can be conducted via least squares, maximum likelihood, and composite likelihood approaches. However, even for univariate cases, likelihood theory is quite cumbersome, especially for higher order models. While multivariate INAR models adequately describe some simple count structures, the models cannot, for example, produce negative correlations in its component series. Moreover, estimation and prediction challenges remain.

5.2. Copula-Based Models

For bivariate and multivariate count series $\mathbf{X}_t = (X_{1,t}, \ldots, X_{d,t})^T$, copulas have been used in several ways. For the bivariate case, Karlis and Pedeli (2013) assume INAR(1) dynamics for the univariate components and use a bivariate copula to induce dependence between Poisson-distributed $E_{1,t}$ and $E_{2,t}$. Likelihood inference can be conducted and the approach is extended to negative binomial series and a copula family having flexible dependencies. Alternatively, one could develop univariate count models for each component series and then use a d-variate copula function to produce the multivariate counts at each time. Heinen and Rengifo (2007) took this approach, employing a Gaussian copula for contemporaneous dependence and specifying univariate models for each marginal series. In this work, likelihood inference is based on adding Uniform$(-1, 0)$ noise to the counts to convert to a continuous setting. Unfortunately, this leads to biased estimators, especially when the counts are small. If likelihood inference is computationally expensive or intractable, then a composite likelihood approach that employs the particular marginal distributions may be feasible. The methods of Heinen and Rengifo (2007) were investigated in a simulation study by Nikoloulopoulos (2013).

The work in Section 3.2 can be extended to the multivariate setting. For example, if the marginal distribution of each component series is specified—say $F_{i,X}$ for the ith component—then a d-dimensional zero-mean unit-variance latent Gaussian process $\{\mathbf{Z}_t\}$ with autocorrelation matrix $\boldsymbol{\rho}(h)$ at lag h can be used to define the count process $X_{i,t} = F_{i,X}^{-1}(\Phi(Z_{i,t}))$. Should one prefer to specify the joint count distributional structure of all d components simultaneously, then multivariate Hermite expansions could be used (Withers 2000). The literature has yet to explore any of these directions.

5.3. Parameter-Driven Models

Recall that parameter-driven models, including those in Section 4, according to the broad categorization in Cox (1981), are models driven by an unobserved process (as opposed to past process values). For multivariate counts, state-space models were studied in Jørgensen et al. (1999) and Jung, Liesenfeld, and Richard (2011); see Ravishanker, Serhiyenko, and Willig (2014) and Ravishanker, Venkatesan, and Hu (2016), among others, for more recent contributions.

We now elaborate on Jørgensen et al. (1999). Consider the case where the conditional distribution of $X_{i,t}$, given an unobserved λ_t, is Poisson with mean $a_{i,t}\lambda_t$, where $a_{i,t}$ can possibly be further modeled. Jørgensen et al. (1999) assumed that $\lambda_0 = 1$ and that λ_t given λ_{t-1} is Gamma distributed with mean $b_t\lambda_{t-1}$

and a squared coefficient of variation of form σ^2/λ_{t-1}, where σ^2 denotes a dispersion parameter and b_t depends on the so-called long-term covariates $\mathbf{c}_t$ through $b_t = \exp(\nabla \mathbf{c}_t^T \boldsymbol{\beta})$, where $\nabla \mathbf{c}_t = \mathbf{c}_t - \mathbf{c}_{t-1}$, with $\mathbf{c}_0 = 0$. The authors show that $\log(\mathbb{E}[\lambda_t]) = \mathbf{c}_t^T \boldsymbol{\beta}$, $\text{var}(\lambda_t) = \phi_t \mathbb{E}[\lambda_t]\sigma^2$, and $\text{cov}(\lambda_t, \lambda_{t+k}) = \phi_t \mathbb{E}[\lambda_{t+k}]\sigma^2$, where $\phi_t = b_t + b_t b_{t-1} + \cdots + b_t b_{t-1} \ldots b_1$. The moment structure of the observed process can be computed. With $\mathbf{a}_t = (a_{1,t}, \ldots, a_{k,t})^T$ and $\boldsymbol{\Lambda}_t = \text{Diag}(a_{1,t}, \ldots, a_{k,t})$, it can be shown that $\mathbb{E}[\mathbf{X}_t] = \mathbf{a}_t \mathbb{E}[\lambda_t]$ and $\text{var}(\mathbf{X}_t) = \boldsymbol{\Lambda}_t \mathbb{E}[\lambda_t] + \mathbf{a}_t \mathbf{a}_t^T \phi_t \mathbb{E}[\lambda_t]\sigma^2$. Hence, the variance matrix of $\mathbf{X}_t$ consists of two components: (1) a "Poisson variance", and (2) an overdispersion component. The authors discussed Kalman prediction and filtering of such processes.

5.4. Observation-Driven Models

Observation driven models are also possible; see Heinen and Rengifo (2007), Liu (2012), Andreassen (2013), Ahmad (2016), and Lee, Lee, and Tjøstheim (2018), which consider linear count models. In a recent contribution, Fokianos et al. (2019) suggested multivariate analogies to Equations (16) and (17) when $p = q = 1$. With a d-dimensional $\{\mathbf{X}_t\}$ and corresponding d-dimensional intensity process $\{\boldsymbol{\lambda}_t\}$, the linear and log-linear models considered are direct extensions of their univariate counterparts.

The linear model is defined by assuming that for each component i,

$$X_{i,t}|\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_{i,t}), \quad \boldsymbol{\lambda}_t = \mathbf{r} + \mathbf{A}\boldsymbol{\lambda}_{t-1} + \mathbf{B}\mathbf{X}_{t-1}, \quad (23)$$

where $\mathbf{r}$ is a d-dimensional with positive entries and $\mathbf{A}$ and $\mathbf{B}$ are $d \times d$ unknown matrices with positive entries (positivity is needed to ensure that $\lambda_{i,t} > 0$ for all i and t). The log-linear model is analogously defined by

$$\begin{aligned} X_{i,t}|\mathcal{F}_{t-1} &\sim \text{Poisson}(\lambda_{i,t}), \\ \boldsymbol{\nu}_t &= \mathbf{r} + \mathbf{A}\boldsymbol{\nu}_{t-1} + \mathbf{B}\log(\mathbf{X}_{t-1} + \mathbf{1}_d), \end{aligned} \quad (24)$$

where $\boldsymbol{\nu}_t = \log(\boldsymbol{\lambda}_t)$ is taken componentwise and $\mathbf{1}_d$ is a d-dimensional vector with all unit entries. For (24), there is no need to impose positivity constraints on $\mathbf{r}$, $\mathbf{A}$, and $\mathbf{B}$. This is an important argument favoring the log-linear model, which accommodates general covariates. The authors introduced a copula function on a vector of associated continuous random variables to describe the data-generating process. This construction avoids joint distribution difficulties (see pages 507 and 508 in Genest and Nešlehová (2007)), but keeps the marginal properties of a Poisson process.

6. Bayesian Dynamic Generalized Linear Models

Bayesian count series modeling approaches have also been actively studied. See Gamerman et al. (2016), Davis et al. (2016), and the therein, for recent Bayesian dynamic generalized linear modeling reviews. Here, count series are developed from a dynamic generalized linear and nonlinear mixed models perspective, with a Bayesian hierarchical modeling (BHM) slant (Holan and Wikle 2016). In this setting, dependence is typically induced via a latent Gaussian process (Cressie and Wikle 2011). Although one can incorporate dependence through non-Gaussian latent processes, non-Gaussian processes are less commonly employed in the literature. We also discuss the modeling approach in Bradley, Holan, and Wikle (2018), which is based on a multivariate log gamma latent process that is computationally attractive.

We now describe a hierarchical paradigm that provides an extremely rich framework for modeling complex dependencies in count series. The main BHM idea is to consider a joint probability model for the data, process, and parameters, which is typically specified through conditionally linked model components; that is, the data conditioned on the process and parameters, and the process conditioned on the parameters. The BHM specification allows complicated marginal dependence structures to be developed via a more scientific specification of conditional means of random process at stages in a model hierarchy. One difference between the models here and in Section 4 lies with the flexibility that can be achieved in the hierarchical structures of this section. While the assumption of conditional independence is often reasonable in many applications, it should be validated whenever possible. See Berliner (1996) and Cressie and Wikle (2011) for broad discussions of this paradigm.

Suppose that $\mathbf{Z} = \{\mathbf{Z}_t\}$ is a latent Gaussian process and that $\mathbf{X} = \{\mathbf{X}_t\}$ denotes data based on $\{\mathbf{Z}_t\}$; for example, $\mathbf{X}_t$ may be a function of $\mathbf{Z}_t$ only as in Section 3.2 or a function of $\mathbf{Z}_t$ and some random error or past process values akin to the models in Section 4. Let $\boldsymbol{\theta}_\mathbf{X}$ denote all parameters arising in $\mathbf{X}$ (these are typically means and covariances) and $\boldsymbol{\theta}_\mathbf{Z}$ denote parameters governing the temporal evolution of $\{\mathbf{Z}_t\}$ (at time t, this structure is assumed to only depend on past, not future process values). For general notation, let $[\mathbf{X}|\mathbf{Z}]$ and $[\mathbf{Z}]$ denote the conditional distribution of $\mathbf{X}$ given $\mathbf{Z}$ and the marginal distribution of $\mathbf{Z}$, respectively. Assuming conditional independence of parameters in $\mathbf{X}$ and $\mathbf{Z}$, the law of total probability shows that the joint probability distribution of the data and process given the parameters factors into

$$[\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}_\mathbf{X}, \boldsymbol{\theta}_\mathbf{Z}] = [\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}_\mathbf{X}][\mathbf{Z}|\boldsymbol{\theta}_\mathbf{Z}], \quad (25)$$

where $[\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}_\mathbf{X}]$ is the data distribution or "data model" (conditional likelihood) and $[\mathbf{Z}|\boldsymbol{\theta}_\mathbf{Z}]$ denotes the latent Gaussian process or "process model."

Our discussion will focus on the process-driven case, where models are developed for $\mathbf{Z}$. This approach has several advantages. First, in many applications, one is interested in predicting values of $\mathbf{Z}$. Models for $\mathbf{Z}$ can directly incorporate scientific insight and simultaneously account for measurement (and/or sampling) and process uncertainties. Second, given the complexity and high-dimensionality of many real-world series, it is often difficult to specify a realistic dependence structure for $\mathbf{X}$ (e.g., due to non-Gaussianity, non-linearity and/or non-stationarity); as a consequence, likelihood-based inference for observation-driven models is often challenging.

Interest typically resides in estimating the posterior distribution of the latent Gaussian process and parameters given the data. Using the Bayes theorem, the BHM obeys

$$[\mathbf{Z}, \boldsymbol{\theta}_\mathbf{Z}, \boldsymbol{\theta}_\mathbf{X}|\mathbf{X}] \propto [\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}_\mathbf{X}][\mathbf{Z}|\boldsymbol{\theta}_\mathbf{Z}][\boldsymbol{\theta}_\mathbf{X}, \boldsymbol{\theta}_\mathbf{Z}], \quad (26)$$

where the normalizing constant integrates over $\mathbf{Z}$, $\boldsymbol{\theta}_\mathbf{X}$, and $\boldsymbol{\theta}_\mathbf{Z}$. An important feature of Equation (26) is that the right hand side can be further decomposed into submodels. For example,

assuming conditional independence given the latent Gaussian process, multiple datasets with different temporal aspects could be accommodated through a data model with specification

$$[\mathbf{X}^{(1)}, \mathbf{X}^{(2)} | \mathbf{Z}, \boldsymbol{\theta}_{\mathbf{X}}^{(1)}, \boldsymbol{\theta}_{\mathbf{X}}^{(2)}] = [\mathbf{X}^{(1)} | \mathbf{Z}, \boldsymbol{\theta}_{\mathbf{X}}^{(1)}][\mathbf{X}^{(2)} | \mathbf{Z}, \boldsymbol{\theta}_{\mathbf{X}}^{(2)}], \quad (27)$$

where, for $j = 1, 2$, $\mathbf{X}^{(j)}$ and $\boldsymbol{\theta}_{\mathbf{X}}^{(j)}$ correspond to the observations and parameters from the jth dataset, respectively (see, e.g., Wang et al. 2012). In this context, $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ need not have the same data distribution. For many applications, it is natural to decompose the process model into subcomponents. For instance, a first-order Markov structure yields the decomposition

$$[(\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_T)] = [\mathbf{Z}_0] \prod_{t=1}^{T} [\mathbf{Z}_t | \mathbf{Z}_{t-1}].$$

Alternatively, the process model could be further decomposed to accommodate multivariate structure. In this case, letting $[\mathbf{Z}] = [\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}]$, the process model can be expressed as $[\mathbf{Z}] = [\mathbf{Z}^{(2)} | \mathbf{Z}^{(1)}][\mathbf{Z}^{(1)}]$, where the order of conditioning is usually suggested by the specific application or chosen by the practitioner (Royle and Berliner 1999).

BHM modeling and inference proceeds by specifying a data model $[\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}_{\mathbf{X}}]$ and a prior distribution $[\boldsymbol{\theta}_{\mathbf{X}}, \boldsymbol{\theta}_{\mathbf{Z}}]$. In the next subsections, these are described for univariate and vector-valued counts, respectively.

6.1. Univariate Count Series Modeling

For univariate count series, the data model $[\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}_{\mathbf{X}}]$ is typically drawn from an exponential family distribution (Poisson, Negative Binomial, Conway-Maxwell Poisson (CMP)); that is,

$$[X_t | \alpha_t] = \exp\{\varphi_t(X_t) + \alpha_t X_t - A_t(\alpha_t)\},$$

where α_t denotes the univariate natural parameter of the exponential family, $\varphi_t(\cdot)$ is a known function, and $A_t(\cdot)$ is a normalizing constant—compare to Equation (9). We assume the usual regularity conditions for exponential families in McCulloch, Searle, and Nehaus (2008),

Our notation uses $\mu_t = \mathbb{E}[X_t|\alpha_t]$. For link functions $g(\cdot)$ and covariates, one often uses the form

$$g(\mu_t) = \mathbf{c}_t^T \boldsymbol{\beta} + Z_t, \quad (28)$$

where $\mathbf{c}_t$ is a vector of known covariates at time t, $\boldsymbol{\beta}$ are unknown regression coefficients that could be further modeled, and $\{Z_t\}$ is IID zero mean Gaussian error (compare to Equation (10)).

For Poisson series, $\{\mu_t\}$ is the intensity function associated with Equation (28) and $g(\cdot)$ is the log link. While the Poisson model implies marginal equidispersion, overdispersion can be achieved by placing a random effect in a latent Gaussian process model for $\{\log(\mu_t)\}$, or, for example, by using a negative binomial (NegBin) data model—say $X_t|\mu_t, \nu \sim$ IID NegBin(Z_t, ν), where $\log(\mu_t)$ can be specified similarly to Equation (28), and ν (or $\log(\nu)$) can be given an appropriate hyperprior. A less common distribution used to model over- or under-dispersed counts is the Conway-Maxwell Poisson (CMP) (Wu, Holan, and Wikle 2013): $X_t|\mu_t, \nu \sim$ IID CMP(Z_t, ν), where $\log(\mu_t)$ can be specified similarly to Equation (28), and $\nu = 1$, $\nu < 1$, and $\nu > 1$ correspond, respectively, to equi-, over-, and under-dispersed distributions.

For certain combinations of intensity and dispersion parameters, calculation of the normalizing constant can be computationally intensive. This may be facilitated using the asymptotic approximations in Minka et al. (2003), or by taking advantage of parallel computing through Open Multiprocessing (OpenMP) and Compute Unified Device Architecture (CUDA); that is,, graphics processing units (Wu, Holan, and Wikle 2013). Distributional specifications are easily generalizable to zero-inflated Poissons (ZIP), or mixture Poissons.

For example, Ravishanker, Venkatesan, and Hu (2016) developed a BHM for a marketing scenario with a ZIP distribution for the prescription counts of a drug, where $X_{i,t}$ is the prescription count in week t by physician i. Let $\lambda_{i,t} = \mathbb{E}[X_{i,t}]$ and $\Pi_{i,t}$ be the corresponding probability of a zero count. The ZIP distribution of $X_{i,t}$ can be decomposed as the mixture $X_{i,t} = V_{i,t}(1 - B_{i,t})$, where $B_{i,t} \sim$ Bernoulli$(\Pi_{i,t})$, $V_{i,t} \sim$ Poisson$(\mu_{i,t})$, and $B_{i,t}$ and $V_{i,t}$ are independent. Exogenous predictors such as the log of the number of sales calls made to physician i during week t (denoted $D_{i,t}$), and the recency-frequency-monetary value (RFM) variable (denoted $R_{i,t}$) are used to model the log and logit, respectively, of the latent physician specific dynamic variables $\mu_{i,t}$ and $\Pi_{i,t}$.

Let $\boldsymbol{\beta}_{i,t}^{\mu} = (\beta_{0,i,t}^{\mu}, \beta_{1,i,t}^{\mu}, \beta_{2,i,t}^{\mu})^T$ and $\boldsymbol{\beta}_{i,t}^{\Pi} = (\beta_{0,i,t}^{\Pi}, \beta_{1,i,t}^{\Pi}, \beta_{2,i,t}^{\Pi})^T$ denote the physician and time specific coefficient vectors, respectively. Including an intercept, a coefficient for $\log(D_{i,t})$ and a coefficient for $R_{i,t}$, each $\boldsymbol{\beta}_{i,t}$ is a three-dimensional vector for each t and i. Then $\boldsymbol{\beta}_{i,t} = ((\boldsymbol{\beta}_{i,t}^{\mu})^T, (\boldsymbol{\beta}_{i,t}^{\Pi})^T)^T$ denotes a $p = 6$-dimensional vector. A hierarchical or structural equation models $\boldsymbol{\beta}_{i,t}$ as a function of a p-dimensional dynamic state vector $\boldsymbol{\gamma}_t$. Physician level variables (such as demographics or specialty, $\mathbf{Z}_i$, if available) and their surveyed attitudes to the prescribed drugs, $\mathbf{C}_{i,t}$, may be included in this equation with static coefficients $\boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}$:

$$\boldsymbol{\beta}_{i,t} = \boldsymbol{\gamma}_t + \mathbf{C}_{i,t}\boldsymbol{\Gamma} + \mathbf{Z}_i\boldsymbol{\Delta} + \mathbf{v}_{i,t}. \quad (29)$$

The errors $\mathbf{v}_{i,t}$ are assumed to be $N_p(\mathbf{0}, \mathbf{V}_i)$ vectors. The state equation describing the dynamic evolution of $\{\boldsymbol{\gamma}_t\}$ is

$$\boldsymbol{\gamma}_t = \mathbf{G}\boldsymbol{\gamma}_{t-1} + \mathbf{w}_t, \quad (30)$$

where $\mathbf{G}$ is the identity matrix if a random walk evolution is assumed, and $\mathbf{w}_t \sim N_p(\mathbf{0}, \mathbf{W})$. Conjugate prior distributions, such as multivariate normal and inverse Wishart, are assumed for the model parameters. Gibbs sampling was used to estimate the posterior distribution of the model parameters. While static coefficients are routinely drawn from known distributions, the forward-filtering-backward-sampling algorithm (Frühwirth-Schnatter 1994) enables sampling $\boldsymbol{\gamma}_t$, and the Metropolis-Hastings algorithm to generate samples of the other parameters. Modeling details, results, and comparisons between several dynamic models are given in Hu (2012), while details and results for the static ZIP models are discussed in Ravishanker, Venkatesan, and Hu (2016).

6.2. Multivariate Modeling

Three approaches will be presented for vector counts. First, Section 6.2.1 describes level correlated models. Next, Section 6.2.2

assumes a particular multivariate data distribution, such as multivariate Poisson, multivariate zero-inflated Poisson (ZIP), or finite mixtures of these distributions. Finally, Section 6.2.3 presents the Poisson multivariate log gamma model.

6.2.1. Level Correlated Models (LCM)

Level correlated models incorporate dependence between components of a count vector $\boldsymbol{X}_{i,t} = (X_{1,i,t}, \ldots, X_{d,i,t})^T$ through an underlying correlated latent process (Aitchison and Ho 1989; Chib and Winklemann 2001; Ma, Kockelman, and Damien 2008). Serhiyenko, Ravishanker, and Venkatesan (2017) describe a marketing application. The observation equation of their LCM is

$$X_{j,i,t}|\boldsymbol{\theta}_{j,i,t} \sim \text{UDC}_j(\boldsymbol{\theta}_{j,i,t}), \tag{31}$$

where UDC stands for "univariate distribution of counts" and $\boldsymbol{\theta}_{j,i,t}$ denotes a set of parameters associated with the UDC. Different valid marginal data distributions may be incorporated for the d component series, such as Poisson, negative binomial, ZIP, and Conway-Maxwell.

The mean $\mu_{j,i,t} = \mathbb{E}[X_{j,i,t}]$ is modeled as a function of predictors via the log link

$$\log(\mu_{j,i,t}) = \beta_{j,i,0} + \gamma_{j,t} + \boldsymbol{c}_{j,i,t}^T \boldsymbol{\beta}_{j,i} + \alpha_{j,i,t}. \tag{32}$$

In Equation (32), the link function for the mean can change with the distribution in Equation (31). For $j = 1, \ldots, d$, the model contains a random subject/location specific intercept $\beta_{j,i,0}$, a time effect for the jth response component, $\gamma_{j,t}$ which is a latent state vector and a random effect $\alpha_{j,i,t}$ which is a response type, time, and subject/location specific level correlated error component. Here, $\boldsymbol{c}_{j,i,t}$ denotes a p_j-dimensional vector of covariates, and $\boldsymbol{\beta}_{j,i}$ is a corresponding p_j-dimensional vector of coefficients, which can be time-varying. Dependence between different components of the count vector can be postulated through $\boldsymbol{\alpha}_{i,t} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i$ is a subject/location specific variance-covariance matrix.

Serhiyenko, Ravishanker, and Venkatesan (2017) used a random walk evolution for $\boldsymbol{\gamma}_t = (\gamma_{1,t}, \ldots, \gamma_{d,t})^T$:

$$\gamma_{j,t} = \gamma_{j,t-1} + w_{j,t}, \tag{33}$$

where $w_{j,t} \sim \text{Normal}(0, 1/W)$; that is, state errors for the different response types are assumed to follow the same normal distribution. Alternately, the state error can follow a response type specific normal distribution; that is, $w_{j,t} \sim \text{Normal}(0, 1/W_j)$. integrated nested Laplace approximations (INLA) Rue, Martino, and Chopin (2009) are used for fast approximate Bayesian modeling using the R-INLA package (r-inla.org).

6.2.2. Hierarchical Multivariate Dynamic Models (HMDM)

We describe a hierarchical model for Poisson counts only; other count distributions can be developed. One definition of a d-variate Poisson random vector $\mathbf{X}$ takes $\mathbf{X} = \mathbf{AZ}$, where $\mathbf{Z} = (Z_1, \ldots, Z_q)^T$ is a q-dimensional ($q \geq d$) vector of unobserved independent Poisson variates, Z_r having mean μ_r, and the components of the $d \times q$ matrix $\mathbf{A}$ take values in $\{0, 1\}$ (Mahamunulu 1967). This construction leaves X_i and X_j as correlated Poisson variates should $a_{i,\ell} = a_{j,\ell} = 1$ for some $\ell \in \{1, \ldots, q\}$. While one cannot build negatively correlated components of $\mathbf{X}$ with such a definition (compare to Section 3.2), the marginal distribution of each component is Poisson. Formally, $\mathbf{X} = (X_1, \ldots, X_d)^T = \mathbf{AZ}$ is said to follow a multivariate Poisson distribution with parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_q)^T$, and PMF

$$\begin{aligned} \text{MP}_d(\mathbf{x}|\boldsymbol{\mu}) &=: \mathbb{P}(\mathbf{X} = \mathbf{x}|\boldsymbol{\mu}) \\ &= \sum_{\mathbf{z} \in g^{-1}(\mathbf{x})} \mathbb{P}(\mathbf{Z} = \mathbf{z}|\boldsymbol{\mu}) \\ &= \sum_{\mathbf{z} \in g^{-1}(\mathbf{x})} \prod_{r=1}^{q} \mathbb{P}(Z_r = z_r|\mu_r). \end{aligned} \tag{34}$$

where $g^{-1}(\cdot)$ denotes the inverse image of the linear function $g(\mathbf{z}) = \mathbf{Az}$ and $\mathbb{P}(Z_r = z_r|\mu_r) = e^{-\mu_r}\mu_r^{z_r}/z_r!$ is the univariate Poisson PMF. The first two moments of $\mathbf{X}$ are $\mathbb{E}[\mathbf{X}|\boldsymbol{\mu}] = \mathbf{A}\boldsymbol{\mu}$ and $\text{var}(\mathbf{X}|\boldsymbol{\mu}) = \mathbf{A}\text{Diag}(\mu_1, \ldots, \mu_q)\mathbf{A}^T$.

The two-way covariance structured multivariate Poisson model proposed in Karlis and Meligkotsidou (2005) used $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2]$, where $\mathbf{A}_1 = \mathbf{I}_d$ captures main effects and $\mathbf{A}_2$ captures two-way covariance effects. Here, $\mathbf{A}_2$ is a $d \times \binom{d}{2}$ binary matrix ($q - d = \binom{d}{2}$) such that each column of $\mathbf{A}_2$ has exactly two ones and $(d-2)$ zeros, and no duplicate columns exist. The parameter $\boldsymbol{\mu}$ is correspondingly partitioned into two parts: $\boldsymbol{\mu}^{(1)} = (\mu_1, \ldots, \mu_d)^T$ contains the main effects and $\boldsymbol{\mu}^{(2)} = (\mu_{d+1}, \ldots, \mu_q)^T$ contains the $\binom{d}{2}$ pairwise covariances $\delta_{r,s} := \text{cov}(Z_r, Z_s)$ for $1 \leq r < s \leq d$.

The probability mass function of a finite mixture of L multivariate Poisson distributions has the representation (Karlis and Meligkotsidou 2007)

$$\mathbb{P}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\pi}) = \sum_{\ell=1}^{L} \pi_\ell \text{MP}_d(\mathbf{x}|\boldsymbol{\mu}_\ell),$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_L)^T$ book keeps mixture proportions. Moments are $\mathbb{E}[\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\pi}] = \sum_{\ell=1}^{L} \pi_\ell \mathbf{A}\boldsymbol{\mu}_\ell$ and

$$\begin{aligned} \text{var}(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\pi}) = \mathbf{A} \Bigg[&\sum_{\ell=1}^{L} \pi_\ell (\text{Diag}(\mu_{1,\ell}, \ldots, \mu_{q,\ell}) + \boldsymbol{\mu}_\ell \boldsymbol{\mu}_\ell^T) \\ &- \left(\sum_{\ell=1}^{L} \pi_\ell \boldsymbol{\mu}_\ell\right)\left(\sum_{\ell=1}^{L} \pi_\ell \boldsymbol{\mu}_\ell\right)^T \Bigg] \mathbf{A}^T. \end{aligned}$$

This structure permits negative correlations in $\mathbf{X}$ should $\boldsymbol{\pi}$ and/or $\boldsymbol{\mu}$ have negatively correlated components (Karlis and Meligkotsidou 2007).

Let $\boldsymbol{X}_{i,t} = (X_{1,i,t}, \ldots, X_{d,i,t})^T$ be observed d-variate count vectors for locations (or subjects) $i = 1, \ldots, n$ at time t. A hierarchical multivariate dynamic model (HMDM) with a d-variate Poisson mixture distribution with two-way covariance structure is now described (Karlis and Meligkotsidou 2007; Ravishanker, Venkatesan, and Hu 2016).

The observation equation is

$$\begin{aligned} \mathbb{P}(\mathbf{x}_{i,t}|\boldsymbol{\mu}_{i,t,\ell}) &= \sum_{\ell=1}^{L} \pi_\ell \text{MP}_d(\mathbf{x}_{i,t}|\boldsymbol{\mu}_{i,t,\ell}), \\ \log(\mu_{r,i,t,\ell}) &= \boldsymbol{B}_{r,i,t}^T \boldsymbol{\delta}_{r,i,t,\ell} + \boldsymbol{S}_{r,i,t}^T \boldsymbol{\eta}_{r,\ell}, r = 1, \ldots, q, \end{aligned} \tag{35}$$

where $\boldsymbol{B}_{r,i,t}$ is an a_r-dimensional vector of exogenous predictors with location-time varying (dynamic) coefficients $\boldsymbol{\delta}_{r,i,t,\ell}$, and $\boldsymbol{S}_{r,i,t}$ is a b_r-dimensional vector of exogenous predictors with static coefficients $\boldsymbol{\eta}_{r,\ell}$. The model is assumed to either include $\delta_{r,i,t,h,1}$, which represents a location-time varying intercept, or $\eta_{r,\ell,1}$, which represents a static intercept with either $D_{r,i,t,1} = 1$ or $S_{r,i,t,1} = 1$. Let $p_d = \sum_{r=1}^{q} a_r$ and $p_s = \sum_{r=1}^{q} b_r$. Let $\boldsymbol{\beta}_{i,t}$ be a p_d-dimensional vector constructed by stacking the a_r coefficients $\boldsymbol{\delta}_{r,i,t}$ for $r = 1, \ldots, q$ into a vector. The structural equation relates the location-time varying parameter $\boldsymbol{\beta}_{i,t}$ to an aggregate (pooled) state parameter $\boldsymbol{\gamma}_t$ via

$$\boldsymbol{\beta}_{i,t} = \boldsymbol{\gamma}_t + \mathbf{A}_{i,t}\boldsymbol{\Gamma} + \mathbf{Z}_i\boldsymbol{\Delta} + \boldsymbol{\nu}_{i,t}, \tag{36}$$

where the errors $\boldsymbol{\nu}_{i,t}$ are assumed IID $N_{p_d}(\mathbf{0}, \boldsymbol{V}_i)$, and $\mathbf{A}_{i,t}$, $\mathbf{Z}_i$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Delta}$ were defined above (29).

The state equation is

$$\boldsymbol{\gamma}_t = \boldsymbol{G}\boldsymbol{\gamma}_{t-1} + \boldsymbol{w}_t, \tag{37}$$

where $\boldsymbol{G}$ is a p_d-dimensional state transition matrix and the state errors $\boldsymbol{w}_t$ are assumed IID $N_{p_d}(\mathbf{0}, \boldsymbol{W})$. When $L = 1$ and $d = 1$, (35)-(37) are a hierarchical DGLM (HDGLM) with the univariate Poisson PMF.

Bayesian inference proceeds under standard prior specifications and uses Gibbs sampling to generate samples from the posterior distribution of the parameters. Details were given in Ravishanker, Venkatesan, and Hu (2016). With small-to-moderate sample sizes, this approach offers a computationally feasible procedure for modeling vector time series of counts. With big data, the approximate LCM approach described in Section 6.2.1 may be preferred.

6.2.3. Multivariate Poisson with Latent Multivariate Log Gamma

This subsection discusses a multivariate conditional Poisson count model based on a multivariate log-gamma latent process (Bradley, Holan, and Wikle 2018). The model was developed for a general multivariate spatio-temporal process; a time series version isolates on the special case of one location. The key advantages to this framework are that the multivariate log gamma distribution is conjugate to the Poisson distribution, which permits model flexibility and fast computational performance.

Let the m-dimensional vector $\mathbf{w} = (w_1, \ldots, w_m)^T$ consist of m mutually independent log gamma random variables, where, for $i = 1, \ldots, m$, we write $w_i \sim \text{LG}(\alpha_i, \kappa_i)$, where α_i and κ_i are positive parameters. Now define $\mathbf{q} = \mathbf{b} + \mathbf{V}\mathbf{w}$, where $\mathbf{V}$ is an $m \times m$ matrix and $\mathbf{b}$ is an m-dimensional vector. We call $\mathbf{q}$ a multivariate log-gamma (MLG) random vector. The PMF of $\mathbf{q}$ was explicitly derived in Bradley, Holan, and Wikle (2018).

Building on Equation (28) and assuming that the data model is Poisson, the Poisson MLG model can be written as

$$g(\boldsymbol{\mu}_t) = \mathbf{c}_t^T\boldsymbol{\beta} + \boldsymbol{\phi}_t^T\boldsymbol{\eta}_t + \boldsymbol{\xi}_t, \tag{38}$$

where $\boldsymbol{\eta}_t$ has a vector autoregressive structure and follows an MLG distribution, $\{\boldsymbol{\phi}_t\}$ is say a deterministic collection of temporal basis functions, $\boldsymbol{\beta}$ follows an MLG distribution, and $\{\boldsymbol{\xi}_t\}$ is independent noise that follows an MLG distribution.

Convenient prior distributions can be placed on the hyperparameters of this model, leading to a fully conjugate model. In this case, performance will improve in situations where the marginal distribution of the latent process is decisively non-Gaussian. Comprehensive details of the theoretical properties and sampling algorithms can be found in Bradley, Holan, and Wikle (2018) and Bradley, Holan, and Wikle (2020).

7. Future Directions

A significant body of current research on thinning methods involves multivariate count series. These methods often combine thinning operators with some of the techniques in this article. Other work looks to extend thinning operators to the set of all integers; Scotto, Weiß, and Gouveia (2015) summarized some operators and approaches for count time series models. Finally, likelihood and composite likelihood estimation are actively being studied

Using binary series to construct different count series shows promise. Since each distributional family requires a different aggregation of the binary series involved, additional research might be needed to achieve the particular marginal distribution sought. For example, Jia, Lund, and Livsey (2020b) easily constructed stationary multinomial distributed series with very flexible autocovariances. However, how to construct stationary generalized Poisson (one of the most flexible count marginal distributions) series with a wide range of autocovariance structures remains unclear. The latent Gaussian transformation techniques in Section 3.2 show vast potential since the resulting series can have any marginal distribution and the most general autocovariances possible. Moreover, they can handle covariates and be fitted via likelihood methods. Here, additional research is needed on two fronts. First, multivariate versions of the methods are needed, including techniques to handle high dimensional settings. Second, while particle filtering-based algorithms can be used to approximate the likelihood (Jia et al. 2020a), these algorithms are only in their developmental infancy and more implementation research is needed to make them efficient.

Analysis of multivariate, and more generally, high-dimensional count series is an area that requires development of more sophisticated statistical methods. Most of the existing theory is based on direct generalizations of univariate models; however, this might not accommodate data observed in different frequencies, series that contain both underdispersed and overdispersed components, or counts with inflated zeroes. Models that take into account such irregularities are hard to develop; in fact, suitable multivariate discrete distributions have yet to be developed. However, copula based approaches, such as those described in Sections 3.2 and 5 provide ways to attack this problem.

Bayesian methods for multivariate count series are computationally demanding in high-dimensional settings. Here, further research is needed on dimension reduction and efficient estimation algorithms. Another direction for future research lies with approximations for Bayesian computation (e.g., variational Bayes and approximate Bayesian computation).

The methods presented in this review article are almost entirely parametric; in general, nonparametric modeling for discrete data is scarce. Recently, there has been an attempt to consider semi-parametric modeling for count series (Zhang 2018).

This allows, for example, in a generalized state-space modeling framework, ways of relaxing a specific parametric PMF (say Poisson or negative binomial) for the observation PMF. The specific PMF would be replaced by an unknown PMF satisfying a log-concave shape constraint. This approach is similar in spirit to that in Chen and Samworth (2013) for continuous-valued series, where the noise distribution is assumed to be log-concave. Background on modeling discrete PMFs using log-concave distributions can be found in Balabdaoui et al. (2013).

Acknowledgments

This article is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not those of the NSF or the U.S. Census Bureau. We thank the referees and Associate Editor for comments that improved this article.

Funding

This research was partially supported by the National Science Foundation (NSF) grants SES-1853096, DMS-1407480, and DMS-1712966, and the Natural Sciences and Engineering Research Council of Canada Discovery Grant 8698.

References

Ahmad, A. (2016), "Contributions à l'Économétrie des Séries Temporelles à Valeurs Entières," Ph.D. thesis, University Charles De Gaulle-Lille III, France. [9]

Ahmad, A., and Franq, C. (2016), "Poisson QMLE of Count Time Series Models," *Journal of Time Series Analysis*, 37, 291–314. [6]

Aitchison, J., and Ho, C. H. (1989), "The Multivariate Poisson-log Normal Distribution," *Biometrika*, 76, 643–653. [11]

Al-Osh, M. A., and Aly, E. E. A. A. (1992), "First Order Autoregressive Time-series With Negative Binomial and Geometric Marginals," *Communications in Statistics — Theory and Methods*, 21, 2483–2492. [2]

Albert, P. S., McFarland, H., Smith, M., and Frank, J. (1994), "Time Series for Modelling Counts From a Relapsing-remitting Disease: Application to Modelling Disease Activity in Multiple Sclerosis," *Statistics in Medicine*, 13, 453–466. [1]

Alqawba, M., Diawara, N., and Rao Chaganty, N. (2019), "Zero-inflated Count Time Series Models Using Gaussian Copula," *Sequential Analysis*, 38, 342–357. [1]

Alzaid, A. A., and Al-Osh, M. A. (1993), "Some Autoregressive Moving Average Processes With Generalized Poisson Marginal Distributions," *Annals of the Institute of Statistical Mathematics*, 45, 223–232. [2]

Andreassen, C. M. (2013), *Models and Inference for Correlated Count Data*," Ph.D. thesis, Aaarhus University, Denmark. [9]

Balabdaoui, F., Jankowski, H., Rufibach, K., and Pavlides, M. (2013), "Asymptotics of the Discrete Log-concave Maximum Likelihood Estimator and Related Applications," *Journal of the Royal Statistical Society*, Series B, 75, 769–790. [13]

Berliner, L. M. (1996), "Hierarchical Bayesian Time Series Models," in *Maximum Entropy and Bayesian Methods*, eds. K. M. Hanson and R. N. Silver, 15–22. Dordrecht, Germany: Springer. [9]

Biller, B. (2009), "Copula-based Multivariate Input Models for Stochastic Simulation," *Operations Research*, **57**, 878–892. [4]

Blight, P. A. (1989), "Time Series Formed From the Superposition of Discrete Renewal Processes," *Journal of Applied Probability*, 26, 189–195. [3,4]

Bollerslev, T. (1986), "Generalized Autoregresive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327. [6]

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2018), "Computationally Efficient Multivariate Spatio-temporal Models for High-dimensional Count-Valued Data" (with discussion), *Bayesian Analysis*, 13, 253–310. [9,12]

——— (2020), "Bayesian Hierarchical Models With Conjugate Full-conditional Distributions for Dependent Data From the Natural Exponential Family," *Journal of the American Statistical Association*, 115, 2037–2052. [12]

Brockwell, P. J. and Davis, R. A. (2016), *Introduction to Time Series and Forecasting* (3rd ed.), New York: Springer-Verlag. [5]

Cario, M. C., and Nelson, B. L. (1997), "Modeling and Generating Random Vectors With Arbitrary Marginal Distributions and Correlation Matrix," Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL. [4]

Chen, H. (2001), "Initialization of NORTA: Generation of Random Vectors With Specified Marginals and Correlations," *INFORMS Journal on Computing*, 13, 312–331. [4]

Chen, Y., and Samworth, R. J. (2013), "Smoothed Log-concave Maximum Likelihood Estimation With Applications," *Statistica Sinica*, 23, 1373–1398. [13]

Chib, S., and Winklemann, R. (2001), "Markov Chain Monte Carlo Analysis of Correlated Count Data," *Journal of Business and Economic Statistics*, 19, 428–435. [11]

Christou, V., and Fokianos, K. (2014), "Quasi-likelihood Inference for Negative Binomial Time Series Models," *Journal of Time Series Analysis*, 35, 55–78. [6]

Cox, D. R. (1981), "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics*, 8, 93–115. [1,8]

Creal, D., Koopman, S. J., and Lucas, A. (2013), "Generalized Autoregressive Score Models With Applications," *Journal of Applied Econometrics*, 28, 777–795. [7]

Cressie, N., and Wikle, C. K. (2011), *Statistics for Spatio-temporal Data*, Hoboken, NJ: Wiley. [9]

Cui, Y., and Lund, R. B. (2009), "A New Look at Time Series of Counts," *Biometrika*, 96, 781–792. [3,4]

Davis, R. A., and Liu, H. (2016), "Theory and Inference for a Class of Nonlinear Models With Application to Time Series of Counts," *Statistica Sinica*, 26, 1673–1707. [6]

Davis, R. A., and Wu, R. (2009), "A Negative Binomial Model for Time Series of Counts," *Biometrika*, 96, 735–749. [5]

Davis, R. A., Wang, Y., and Dunsmuir, W. T. M. (1999), "Modelling Time Series of Count Data," in *Asymptotics, Nonparametric & Time Series*, S. Ghosh, ed., 63–114, New York City: Marcel Dekker. [1,6,7]

Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (2000), "On Autocorrelation in a Poisson Regression Model," *Biometrika*, 87, 491–505. [1,5]

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2003), "Observation-driven Models for Poisson Counts," *Biometrika*, 90, 777–790. [1,7]

——— (2005), "Maximum Likelihood Estimation for an Observation Driven Model for Poisson Counts," *Methodology and Computing in Applied Probability*, 7, 149–159. [7]

Davis, R. A., Holan, S. H., Lund, R. B., and Ravishanker, N., eds. (2016), *Handbook of Discrete-Valued Time Series*, Boca Raton, FL: CRC Press. [9]

Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Reviews*, 41, 45–7. [6]

Douc, R., Doukhan, P., and Moulines, E. (2013), "Ergodicity of Observation-driven Time Series Models and Consistency of the Maximum Likelihood Estimator," *Stochastic Processes and their Applications*, 123, 2620–2647. [6]

Douc, R., Moulines, E., and Stoffer, D. S. (2014), *Nonlinear Time Series: Theory, Methods, and Applications with R Examples*, Boca Raton, FL: Chapman & Hall/CRC Press. [8]

Doukhan, P., Fokianos, K., and Tjøstheim, D. (2012), "On Weak Dependence Conditions for Poisson Autoregressions," *Statistics & Probability Letters*, 82, 942–948. With a Correction in Vol. 83, pp. 1926-1927. [6]

Du, J.-G., and Li, Y. (1991), "The Integer-valued Autoregressive (INAR(p)) Model," *Journal of Time Series Analysis*, 12, 129–142. [2,3]

Dunsmuir, W. T. M. (2016), "Generalized Linear Autoregressive Moving Average Models," in *Handbook of Discrete-Valued Time Series*, eds. R. A. Davis, S. H. Holan, R. B. Lund, and N. Ravishanker, 51–76. Boca Raton, FL: CRC Press. [7]

Dunsmuir, W. T. M., Scott, D. J., et al. (2015), "The GLARMA Package for Observation-driven Time Series Regression of Counts," *Journal of Statistical Software*, 67, 1–36. [7]

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007. [6]

Escarela, G., Mena, R. H., and Castillo-Morales, A. (2006), "A Flexible Class of Parametric Transition Regression Models Based on Copulas: Application to Poliomyelitis Incidence," *Statistical Methods in Medical Research*, 15, 593–609. [4]

Fahrmeir, L. and Tutz, G. (2001), *Multivariate Statistical Modelling Based on Generalized Linear Models* (2nd ed.), New York City: Springer. [1]

Ferland, R., Latour, A., and Oraichi, D. (2006), "Integer–valued GARCH Processes," *Journal of Time Series Analysis Analysis*, 27, 923–942. [6]

Fokianos, K. (2012), "Count Time Series Models," in *Handbook of Statistics: Time Series Analysis–Methods and Applications* (vol. 30), eds. T. S. Rao, S. S. Rao, and C. R. Rao, 315–347. Amsterdam: Elsevier B. V. [6]

——— (2015), "Statistical Analysis of Count Time Series Models: A GLM Perspective," In *Handbook of Discrete-Valued Time Series*, eds. R. Davis, S. Holan, R. Lund, and N. Ravishanker, 3–28. Boca Raton, FL: CRC Press. [6]

Fokianos, K., and Kedem, B. (2004), "Partial Likelihood Inference for Time Series Following Generalized Linear Models," *Journal of Time Series Analysis*, 25, 173–197. [1]

Fokianos, K., and Tjøstheim, D. (2011), "Log-linear Poisson Autoregression," *Journal of Multivariate Analysis*, 102, 563–578. [6]

——— (2012), "Nonlinear Poisson Autoregression," *Annals of the Institute of Statistical Mathematics*, 64, 1205–1225. [6]

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009), "Poisson Autoregression," *Journal of the American Statistical Association*, 104, 1430–1439. [6]

Fokianos, K., Støve, B., Tjøstheim, D., and Doukhan, P. (2019), "Multivariate Count Autoregressions," *Bernoulli*, 26, 471–499. [7,9]

Fralix, B., Livsey, J., and Lund, R. (2012), "Renewal Sequences with Periodic Dynamics," *Probability in the Engineering and Informational Sciences*, 26, 1–15. [4]

Franke, J., and Rao, T. S. (1995), "Multivariate First-order Integer Valued Autoregressions," Technical report, Department of Mathematics, UMIST. [8]

Frühwirth-Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, **15**, 183–202. [10]

Gamerman, D., Abanto-Valle, C., Silva, R., and Martins, T. (2016), "Dynamic Bayesian Models for Discrete-valued Time Series," in *Handbook of Discrete-Valued Time Series*, eds. R. A. Davis, S. H. Holan, R. B. Lund, and N. Ravishanker, 165–186. Boca Raton, FL : CRC Press. [9]

Gauthier, G. and Latour, A. (1994), "Convergence Forte des Estimateurs des Paramètres d'un Processus GENAR(*p*)," *Annals of Science, Mathematiques du Québec*, 18, 49–71. [2]

Genest, C., and Nešlehová, J. (2007), "A Primer on Copulas for Count Data," *ASTIN Bulletin*, 37, 475–515. [9]

Graziadei, H., Lopes, H. F., and Marques, P. C. (2020), "A Bayesian inar(1) Model With Adaptive Overdispersion," preprint. [2]

Grigoriu, M. (2007), "Multivariate Distributions With Specified Marginals: Applications to Wind Engineering," *Journal of Engineering Mechanics*, 133, 174–184. [4]

Han, Z., and De Oliveira, V. (2016), "On the Correlation Structure of Gaussian Copula Models for Geostatistical Count Data," *Australian & New Zealand Journal of Statistics*, 58, 47–69. [4]

Harvey, A. C. (2013), *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*, Cambridge, UK: Cambridge University Press. [7]

Heinen, A. (2003), "Modelling Time Series Count Data: An Autoregressive Conditional Poisson Model," Technical Report MPRA Paper 8113, University Library of Munich, Germany. Available at *http://mpra.ub.uni-muenchen.de/8113/*. [6]

Heinen, A., and Rengifo, E. (2007), "Multivariate Autoregressive Modeling of Time Series Count Data Using Copulas," *Journal of Empirical Finance*, 14, 564 – 583. [8,9]

Holan, S. H., and Wikle, C. K. (2016), "Hierarchical Dynamic Generalized Linear Mixed Models for Discrete-valued Spatio-temporal Data," in *Handbook of Discrete-Valued Time Series*, eds. R. A. Davis, S. H. Holan, R. B. Lund, and N. Ravishanker, 327–348. Boca Raton, FL: CRC Press. [9]

Holden, R. T. (1987), "Time Series Analysis of Contagious Process," *Journal of the American Statistical Association*, 82, 1019–1026. [1]

Hu, S. (2012), "Dynamic Modeling of Discrete-valued Time Series With Applications," Ph.D. thesis, University of Connecticut, Storrs, CT, USA. [10]

Jia, Y., Kechagias, S., Livsey, J., Lund, R. B., and Pipiras, V. (2020a), "Latent Gaussian Count Time Series Modeling," *Journal of the American Statistical Association*, in revision. [4,8,12]

Jia, Y., Lund, R. B., and Livsey, J. (2020b), "Superpositioned Stationary Count Time Series," *Probability in the Engineering and Informational Sciences*, to appear. [3,4,12]

Joe, H. (1996), "Time Series Models With Univariate Margins in the Convolution-closed Infinitely Divisible Class," *Journal of Applied Probability*, 33, 664–677. [2]

——— (1997), *Multivariate Models and Dependence Concepts*, London: Chapman & Hall. [2,4,7,8]

——— (2014), *Dependence Modeling With Copulas*, Boca Raton, FL: Chapman & Hall/CRC. [4]

——— (2016), "Markov Count Time Series Models With Covariates," in *Handbook of Discrete-valued Time Series*, eds. R. A. Davis, S. H. Holan, R. B. Lund, and N. Ravishanker, 29–49. Boca Raton, FL : CRC Press. [2,3,7]

——— (2019), "Likelihood Inference for Generalized Integer Autoregressive Time Series Models," *Econometrics*, 7, doi.org/10.3390/econometrics7040043. [3]

Jørgensen, B., Lundbye-Christensen, S., Song, P. X.-K., and Sun, L. (1999), "A State Space Model for Multivariate Longitudinal Count Data," *Biometrika*, 86, 169–181. [8]

Jung, R., Liesenfeld, R., and Richard, J.-F. (2011), "Dynamic Factor Models for Multivariate Count Data: An Application to Stock–market Trading Activity," *Journal of Business & Economic Statistics*, 29, 73–85. [8]

Karlis, D. (2016), "Modelling Multivariate Times Series for Counts," In *Handbook of Discrete-Valued Time Series*, eds. R. A. Davis, S. H. Holan, R. B. Lund, and N. Ravishanker, 407–424. Boca Raton, FL: CRC Press. [8]

Karlis, D., and Meligkotsidou, L. (2005), "Multivariate Poisson Regression With Covariance Structure," *Statistics and Computing*, 15, 255–265. [11]

——— (2007), "Finite Mixtures of Multivariate Poisson Regression With Application," *Journal of Statistical Planning and Inference*, 137, 1942–1960. [11]

Karlis, D., and Pedeli, X. (2013), "Flexible Bivariate INAR(1) Processes Using Copulas," *Communications in Statistics — Theory and Methods*, 42, 723–740. [8]

Kedem, B. (1980), "Estimation of the Parameters in Stationary Autoregressive Processes After Hard Limiting," *Journal of the American Statistical Association*, 75, 146–153. [3]

Kedem, B., and Fokianos, K. (2002), *Regression Models for Time Series Analysis*, Hoboken, NJ: Wiley. [1]

Latour, A. (1997), "The Multivariate GINAR(p) Process," *Advances in Applied Probability*, 29, 228–248. [8]

——— (1998), "Existence and Stochastic Structure of a Non-negative Integer-valued Autoregressive Process," *Journal of Time Series Analysis*, 19, 439–455. [2]

Lee, Y., Lee, S., and Tjøstheim, D. (2018), "Asymptotic Normality and Parameter Change Test for Bivariate Poisson INGARCH Models," *TEST*, 27, 52–69. [9]

Li, W. K. (1994), "Time Series Models Based on Generalized Linear Models: Some Further Results," *Biometrics*, 50, 506–511. [1]

Liboschik, T., Fokianos, K., and Fried, R. (2017), "tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models," *Journal of Statistical Software*, 82. [6]

Liu, H. (2012), "Some Models for Time Series of Counts," Ph.D. thesis, Columbia University, USA. [9]

Livsey, J., Lund, R. B., Kechagias, S., and Pipiras, V. (2018), "Multivariate Integer-valued Time Series With Flexible Autocovariances and Their Application to Major Hurricane Counts," *Annals of Applied Statistics*, 12, 408–431. [3,4,8]

Lund, R. B. and Livsey, J. (2016). Renewal-based count time series. In R. A. Davis, S. H. Holan, R. B. Lund, and N. Ravishanker, editors, *Handbook of Discrete-Valued Time Series*, 101–120. Boca Raton, FL: CRC Press. [4]

Lund, R. B., Holan, S. H., and Livsey, J. (2016), "Long Memory Discrete-Valued Time Series," in R. A. Davis, S. H. Holan, R. B. Lund, and N. Ravishanker, editors, *Handbook of Discrete-Valued Time Series*, 447–458. CRC Press, Boca Raton, Florida. [3,4]

Ma, J., Kockelman, K. M., and Damien, P. (2008), "A Multivariate Poisson-lognormal Regression Model for Prediction of Crash Counts by Severity, Using Bayesian Methods," *Accident Analysis and Prevention*, 40, 964–975. [11]

Mahamunulu, D. (1967), "A Note on Regression in the Multivariate Poisson Distribution," *Journal of the American Statistical Association*, 62, 251–258. [11]

Masarotto, G., and Varin, C. (2012), "Gaussian Copula Marginal Regression," *Electronic Journal of Statistics*, 6, 1517–1549. [4]

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman & Hall. [1,6]

McCulloch, C. E., Searle, S. R., and Nehaus, J. M. (2008), "Generalized, Linear, and Mixed Models," 2nd Edition, Hoboken, NJ: Wiley. [10]

McKenzie, E. (1986), "Autoregressive Moving-average Processes With Negative-binomial and Geometric Marginal Distributions," *Advances in Applied Probability*, 18, 679–705. [2]

McKenzie, E. (1988), "Some ARMA Models for Dependent Sequences of Poisson Counts," *Advances in Applied Probability*, 20, 822–835. [2]

Minka, T. P., Shmueli, G., Kadane, J. B., Borle, S., and Boatwright, P. (2003), "Computing With the COM-Poisson Distribution," Technical report. [10]

Nelder, J. A., and Wedderburn, R. W. M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society*, Series A, 135, 370–384. [1]

Neumann, M. (2011), "Absolute Regularity and Ergodicity of Poisson Count Processes," *Bernoulli*, 17, 1268–1284. [6]

Nikoloulopoulos, A. K. (2013), "On the Estimation of Normal Copula Discrete Regression Models Using the Continuous Extension and Simulated Likelihood," *Journal of Statistical Planning and Inference*, 143, 1923–1937. [8]

Paul, M., Held, L., and Toschke, A. M. (2008), "Multivariate Modelling of Infectious Disease Surveillance Data," *Statistics in Medicine*, 27, 6250–6267. [8]

Pedeli, X., and Karlis, D. (2013a), "On Composite Likelihood Estimation of a Multivariate INAR(1) Model," *Journal of Time Series Analysis*, 34, 206–220. [2,8]

——— (2013b), "Some Properties of Multivariate INAR(1) Processes," *Computational Statistics & Data Analysis*, 67, 213 – 225. [8]

Pipiras, V., and Taqqu, M. S. (2017), *Long-Range Dependence and Self-Similarity*, Cambridge, UK: Cambridge University Press. [4]

Ravishanker, N., Serhiyenko, V., and Willig, M. R. (2014), "Hierarchical Dynamic Models for Multivariate Times Series of Counts," *Statistics and Its Interface*, 7, 559–570. [8]

Ravishanker, N., Venkatesan, R., and Hu, S. (2016), "Dynamic Models for Time Series of Counts With a Marketing Application," in *Handbook of Discrete-Valued Time Series*, eds. R. A. Davis, S. H. Holan, R. B. Lund, and N. Ravishanker, 425–446. Boca Raton, FL : CRC Press. [8,10,11,12]

Royle, J. A., and Berliner, L. M. (1999), "A Hierarchical Approach to Multivariate Spatial Modeling and Prediction," *Journal of Agricultural, Biological, and Environmental Statistics*, 4, 29–56. [10]

Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations," *Journal of the Royal Statistical Society*, Series B, 1, 319–392. [11]

Rydberg, T. H. and Shephard, N. (2000), "A Modeling Framework for the Prices and Times of Trades on the New York Stock Exchange," In *Nonlinear and Nonstationary Signal Processing*, eds. W. J. Fitzgerlad, R. L. Smith, A. T. Walden, and P. C. Young, 217–246. Cambridge: Isaac Newton Institute and Cambridge University Press. [6]

Scotto, M. G., Weiß, C. H., Silva, M. E., and Pereria, I. (2014), "Bivariate Binomial Autoregressive Models," *Journal of Multivariate Analysis*, 125, 233–251. [8]

Scotto, M. G., Weiß, C. H., and Gouveia, S. (2015), "Thinning-based Models in the Analysis of Integer-valued Time Series: A Review," *Statistical Modelling*, 15, 590–618. [2,12]

Sellers, K. F., Peng, S. J., and Arab, A. (2020), "A Flexible Univariate Autoregressive Time-series Model for Dispersed Count Data," *Journal of Time Series Analysis*, 41, 436–453. [1]

Serhiyenko, V., Ravishanker, N., and Venkatesan, R. (2017), "Multi-stage Multivariate Modeling of Temporal Patterns in Prescription Counts for Competing Drugs in a Therapeutic Category" (with discussion), *Applied Stochastic Models in Business and Industry*, 34, 61–78. [11]

Sim, T. (2016), "Maximum Likelihood Estimation in Partially Observed Markov Models With Applications to Time Series of Counts," Ph.D. thesis, Télécom ParisTech, Paris. [6]

Smith, M. S., and Khaled, M. A. (2012), "Estimation of Copula Models With Discrete Margins Via Bayesian Data Augmentation," *Journal of the American Statistical Association*, 107, 290–303. [4]

Streett, S. (2000), "Some Observation Driven Models for Time Series of Counts," Ph.D. thesis, Colorado State University, Department of Statistics, Fort Collins, CO. [6]

Tong, H. (1990a), *Non-Linear Time Series. A Dynamical System Approach*, New York City: Oxford University Press. [7]

Tong, Y. L. (1990b), *The Multivariate Normal Distribution*, New York City: Springer-Verlag. [3]

Wang, C., Liu, H., Yao, J.-F., Davis, R. A., and Li, W. K. (2014), "Self-excited Threshold Poisson Autoregression," *Journal of the American Statistical Association*, 109, 777–787. [6,7]

Wang, J. C., Holan, S. H., Nandram, B., Barboza, W., and Toto, C., and Anderson, E. (2012), "A Bayesian Approach to Estimating Agricultural Yield Based on Multiple Repeated Surveys," *Journal of Agricultural, Biological, and Environmental Statistics*, 17, 84–106. [10]

Weiß, C. H. (2008), "Serial Dependence and Regression of Poisson INARMA Models," *Journal of Statistical Planning and Inference*, 138, 2975–2990. [2]

——— (2018), *An Introduction to Discrete-Valued Time Series*, Hoboken, NJ: Wiley. [2]

Whitt, W. (1976), "Bivariate Distributions With Given Marginals," *The Annals of Statistics*, 4, 1280–1289. [4]

Withers, C. S. (2000), "A Simple Expression for Multivariate Hermite Polynomials," *Statistics and Probability Letters*, pages 165–169. [8]

Wong, W. H. (1986), "Theory of Partial Likelihood," *Annals of Statistics*, 14, 88–123. [1]

Wu, G., Holan, S. H., and Wikle, C. K. (2013), "Hierarchical Bayesian Spatio-temporal Conway–Maxwell Poisson Models With Dynamic Dispersion," *Journal of Agricultural, Biological, and Environmental Statistics*, 18, 1–22. [10]

Wu, W., and Shao, X. (2004), "Limit Theorems for Iterated Random Functions," *Journal of Applied Probability*, 41, 425–436. [6]

Zeger, S. L. (1988), "A Regression Model for Time Series of Counts," *Biometrika*, 75, 621–629. [1,5]

Zeger, S. L., and Qaqish, B. (1988), "Markov Regression Models for Time Series: A Quasi-likelihood Approach," *Biometrics*, 44, 1019–1031. [1,6]

Zhang, J. (2018), "Time Series Modeling With Shape Constraints," Ph.D. thesis, Columbia University, Department of Statistics. [5,12]

Zhu, R., and Joe, H. (2010), "Count Data Time Series Models Based on Expectation Thinning," *Stochastic Models*, 26, 431–462. [3]

Zucchini, W., and MacDonald, I. L. (2009), *Hidden Markov Models for Time Series: An Introduction Using R*, Boca Raton, FL: Chapman & Hall/CRC. [8]

——— (2016), "Hidden Markov Models for Discrete-valued Time Series," *Handbook of Discrete-Valued Time Series*, eds. R. A. Davis, S. H. Holan, R. B. Lund, and N. Ravishanker, 267–286. Boca Raton, FL: CRC Press. [7]