19

Models for Multivariate Count Time Series

Dimitris Karlis

CONTENTS

19.1 Introduction...................................................................................407
19.2 Models for Multivariate Count Data.......................................................408
  19.2.1 Use of Multivariate Extensions of Simple Models...............................408
  19.2.2 Models Based on Copulas...........................................................409
  19.2.3 Other Multivariate Models for Counts............................................411
19.3 Models Based on Thinning..................................................................411
  19.3.1 The Standard INAR Model..........................................................411
  19.3.2 Multivariate INAR Model...........................................................412
  19.3.3 Estimation.............................................................................415
  19.3.4 Other Models in This Category.....................................................417
19.4 Parameter-Driven Models...................................................................418
  19.4.1 Latent Factor Model..................................................................418
  19.4.2 State Space Model....................................................................418
19.5 Observation-Driven Models.................................................................419
19.6 More Models...................................................................................421
19.7 Discussion......................................................................................421
References............................................................................................422

19.1 Introduction

We have seen a tremendous increase in models for discrete-valued time series over the past few decades. Although there is a flourishing literature on models and methods for univariate integer-valued time series, the literature is rather sparse for the multivariate case, especially for multivariate count time series. Multivariate count data occur in several different disciplines like epidemiology, marketing, criminology, and engineering, just to name a few. For example, in syndromic surveillance systems, we record the number of patients with a given symptom. An abrupt change in this number could indicate a threat to public health, and our goal would be to discover such a change as early as possible. In practice, a large number of symptoms are counted creating possibly associated multiple time series of counts. An adequate analysis of such multiple time series requires models that can take into account the correlation across time as well as the correlations between the different symptoms.
Another example comes from geophysical research, where data are collected over time on the number of earthquakes whose magnitudes are above a certain threshold (Boudreault and Charpentier, 2011). Different series can be generated from adjacent areas, making an important scientific question the correlation between the two areas. In criminology, one counts the number of occurrences of one type of crime in successive time periods (say, weeks). Analyzing together more than one type of crime generates many count time series that may be correlated. In finance, an analyst might wish to model the number of bids and asks for a stock, or the number of trades of different stocks in a portfolio. Similar examples may be seen for the number of purchases of different but related products in marketing, the number of claims of different policies in actuarial science, etc.

The underlying similarity in all the earlier mentioned examples is that the collected data are correlated counts observed at different time points. Hence, we have two sources of correlation, serial correlation since the data are time series and cross-correlation since the component time series are correlated at each time point. The need to account for both serial and cross-correlation complicates model specification, estimation, and inference. The literature on statistical models for multivariate time series of counts is rather sparse, perhaps because the analytical and computational issues are complicated. In recent years, new models have been developed to facilitate the modeling approach, which we discuss in this chapter.

We start in Section 19.2 with a brief review of some models for multivariate count data and a discussion of the problems that arise. These models form the basis for the time series models that will be discussed in the following sections along three main avenues: models based on thinning (Section 19.3), parameter-driven models (Section 19.4), and observation-driven models (Section 19.5). Additional models will be mentioned in Section 19.6. Concluding remarks are given in Section 19.7.

### 19.2 Models for Multivariate Count Data

Even ignoring the time correlation there are not many models for multivariate counts in the literature. Inference for multivariate counts is analytically and computationally demanding. Perhaps the case is easier and more developed in the bivariate case but there are several bivariate models that cannot easily generalize to the multivariate. This is a major obstacle for the development of flexible models to be used also in the time series context. We briefly explore some of the issues.

#### 19.2.1 Use of Multivariate Extensions of Simple Models

Consider, for example, the simplest extension of the univariate Poisson distribution to the bivariate case. As in Kocherlakota and Kocherlakota (1992), the bivariate Poisson has probability mass function (pmf)

\[
P(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2; \Theta) = e^{-\theta_1 - \theta_2 - \theta_0} \frac{\theta_1^{y_1}}{y_1!} \frac{\theta_2^{y_2}}{y_2!} \sum_{s=0}^{\min(y_1, y_2)} \binom{y_1}{s} \binom{y_2}{s} s! \left( \frac{\theta_0}{\theta_1 \theta_2} \right)^s, \tag{19.1}
\]
where $\theta_1, \theta_2, \theta_0 \geq 0$, $y_1, y_2 = 0, 1, \ldots$, $\Theta = (\theta_1, \theta_2, \theta_0)$. $\theta_0$ is the covariance while the marginal means and variances are equal to $\theta_1 + \theta_0$ and $\theta_2 + \theta_0$, respectively. The marginal distributions are Poisson. One can easily see that this pmf involves a finite summation that can be computationally intensive for large counts. This bivariate Poisson distribution only allows positive correlation. We denote this by $BP(\theta_1, \theta_2, \theta_0)$. For $\theta_0 = 0$, we get two independent Poisson distributions. We may generalize this model by considering mixtures of the bivariate Poisson. Although there are a few schemes, two ways to do this have been studied in detail. Most of the literature assumes a $BP(\alpha \theta_1, \alpha \theta_2, \alpha \theta_0)$ distribution and places a mixing distribution on $\alpha$. Depending on the choice of the distribution of $\alpha$, such a model produces overdispersed marginal distributions but with always positive correlation. The correlation comes from two sources, the first is the intrinsic one from $\theta_0$ and the second is due to the use of a common $\alpha$.

A more refined model can be produced by assuming a $BP(\theta_1, \theta_2, 0)$ and letting $\theta_1, \theta_2$ jointly vary according to some bivariate continuous distribution, as, for example, in Chib and Winkelmann (2001) where a bivariate lognormal distribution is assumed. Here, the correlation comes from the correlation of the joint mixing distribution, and thus, it can be negative as well. The obstacle is that we do not have flexible bivariate distributions to use for the mixing, or some of them may lead to computational problems. The bivariate Poisson lognormal distribution in Chib and Winkelmann (2001) does not have closed-form pmf and bivariate integration is needed.

It is interesting to point out that generalization to higher dimensions is not straightforward even for simple models. For example, generalizing the bivariate Poisson to the multivariate Poisson with one correlation parameter for every pair of variables leads to multiple summation, see the details in Karlis and Meligkotsidou (2005). We will see later some ideas on how to overcome these problems.

### 19.2.2 Models Based on Copulas

A different avenue to build multivariate models is to apply the copula approach. Copulas (see Nelsen, 2006) have found a remarkably large number of applications in finance, hydrology, biostatistics, etc., since they allow the derivation and application of flexible multivariate models with given marginal distributions. The key idea is that the marginal properties can be separated from the association properties, thus leading to a wealth of potential models. For the case of discrete data, copula-based modeling is less developed. Genest and Nešlehová (2007) provided an excellent review on the topic. It is important to keep in mind that some of the desirable properties of copulas are not valid when dealing with count data. For example, dependence properties cannot be fully separated from marginal properties. To see this, consider the Kendall’s tau correlation coefficient. The probability for a tie is not zero for discrete data and depends on the marginal distribution, hence the value of Kendall’s tau is also dependent on the marginal distributions. Furthermore, the pmf cannot be derived through derivatives but via finite differences which can be cumbersome in larger dimensions. For a recent review on copulas for discrete data, see Nikoloulopoulos (2013b).

To help the exposition we first discuss bivariate copulas.

**Definition (Nelsen, 2006).** A bivariate copula is a function $C$ from $[0,1]^2$ to $[0,1]$ with the following properties: (a) for every $(u,v) \in [0,1]$, $C(u,0) = 0 = C(0,v)$ and $C(u,1) = u$, ...
That is, copulas are bivariate distributions with uniform marginals. Recall the inversion theorem, central in simulation, where starting from a uniform random variable and applying the inverse transform of a distribution function we can generate any desired distribution. Copulas extend this idea in the sense that we start from two correlated uniforms and hence we end up with variables from whatever distribution we like which are still correlated.

If \( F(x) \) and \( G(y) \) are the cdfs of the univariate random variables \( X \) and \( Y \), then \( C(F(x), G(y)) \) is a bivariate distribution for \( (X, Y) \) with marginal distributions \( F \) and \( G \), respectively. Conversely, if \( H \) is a bivariate cdf with univariate marginal cdfs \( F \) and \( G \), then, according to Sklar's theorem (Sklar, 1959) there exists a bivariate copula \( C \) such that for all \((x, y)\), \( H(x, y) = C(F(x), G(y)) \). If \( F \) and \( G \) are continuous, then \( C \) is unique, otherwise, \( C \) is uniquely determined on \( \text{range } F \times \text{range } G \). This lack of uniqueness is not a problem in practical applications as it implies that there may exist two copulas with identical properties.

Copulas provide the joint cumulative function. In order to derive the joint density (for continuous data) or the joint probability function (for discrete data) we need to take the derivatives or the finite differences of the copula. For bivariate discrete data, the pmf is obtained by finite differences of the cdf through its copula representation (Genest and Nešlehová, 2007), that is,

\[
h(x, y; \alpha_1, \alpha_2, \theta) = C(F(x; \alpha_1), G(y; \alpha_2); \theta) - C(F(x - 1; \alpha_1), G(y; \alpha_2); \theta)
- C(F(x; \alpha_1), G(y - 1; \alpha_2); \theta) + C(F(x - 1; \alpha_1), G(y - 1; \alpha_2); \theta),
\]

where \( F(\cdot) \) and \( G(\cdot) \) are the marginal cdfs and \( \alpha_1 \) and \( \alpha_2 \) are the parameters associated with the respective marginal distributions and \( \theta \) denotes the parameter(s) of the copula. This poses a big problem in higher dimensions. In order to take differences, we need to evaluate the copula eight times in the trivariate case and \( 2^d \) times for \( d \) dimensions.

Copulas are cdfs and thus in many cases they are given as multidimensional integrals and not as simple formulas. A simple example is the bivariate Gaussian copula which is defined as a bivariate integral. In this case, one needs to evaluate multidimensional integrals many times in order to evaluate the pmf. To avoid this extensive integration, one can switch to copulas that are given in simple form without the need to integrate (e.g., the Frank copula). But even in this case, one needs to add and subtract several numbers (which are usually very close to 0) leading to possible truncation errors.

Another problem relates to the shortage of copulas that can allow for flexible correlation structure. For example, the multivariate Archimedean copulas assign the same correlation to all pairs of variables, which is too restrictive in practice. Also if one needs to specify both positive and negative correlations, more restrictions apply. To sum up, a big issue in working with models defined via copulas is the lack of a framework that allows flexible structure while maintaining computational simplicity.
19.2.3 Other Multivariate Models for Counts

There are other strategies to build flexible models for multivariate counts such as models based on conditional distributions (Berkhout and Plug, 2004), or finite mixtures (Karlis and Meligkotsidou, 2007). In most cases, things are more complicated than continuous models where the multivariate normal distributions is a cornerstone allowing for great flexibility and feasible calculations.

## 19.3 Models Based on Thinning

### 19.3.1 The Standard INAR Model

Among the most successful integer-valued time series models proposed in the literature are the INteger-valued AutoRegressive (INAR) models, introduced by McKenzie (1985) and Al-Osh and Alzaid (1987). Since then, several articles have been published to extend and generalize these models. The reader is referred to McKenzie (2003) and Jung and Tremayne (2006) for a comprehensive review of such models. The extension of the simple INAR(1) process to the multidimensional case is interesting as it provides a general framework for multivariate count time series modeling. The model had been considered in Franke and Rao (1995) and Latour (1997) but since then, there has been a long hiatus since this topic was addressed again by Pedeli and Karlis (2011) and Boudreault and Charpentier (2011).

**Definition** A sequence of random variables \( \{Y_t : t = 0, 1, \ldots \} \) is an INAR(1) process if it satisfies a difference equation of the form

\[
Y_t = \alpha \circ Y_{t-1} + R_t; \quad t = 1, 2, \ldots,
\]

where \( \alpha \in [0, 1] \), \( R_t \) is a sequence of uncorrelated nonnegative integer-valued random variables with mean \( \mu \) and finite variance \( \sigma^2 \) (called hereafter as the innovations), and \( Y_0 \) represents an initial value of the process.

The operator “\( \circ \)” is defined as

\[
\alpha \circ Y = \sum_{i=1}^{Y} Z_i = Z,
\]

(19.2)

where \( Z_i \) are independently and identically distributed Bernoulli random variables with \( P(Z_i = 1) = 1 - P(Z_i = 0) = \alpha \). This operator, known as the binomial thinning operator, is due to Steutel and van Harn (1979) and mimics the scalar multiplication used for normal time series models so as to ensure that only integer values will occur.
Note that by assuming any distribution other than Bernoulli for \( Z_i \)s in (19.2), we get a generalized Steutel and van Harn operator. Other operators can also be similarly defined. A review on thinning operators can be found in Weiß (2008).

The model can be generalized to have \( p \) terms (i.e., INAR(\( p \))), but there is no unique way to do this. Moving average (MA) terms can be added to the model leading to INARMA models. Covariates can also be introduced to model the mean of the innovation term to allow measuring the effect of additional information leading to INAR regression models.

We next present extensions to the multivariate case by first extending the thinning operator to a matrix-valued form and then presenting the multivariate INAR model.

Let \( A \) be an \( r \times r \) matrix with elements \( \alpha_{ij}, i, j = 1, \ldots, r \) and \( Y \) be a nonnegative integer-valued \( r \)-dimensional vector. The matrix-valued operator “\( \circ \)” is defined as

\[
A \circ Y = \begin{pmatrix}
\sum_{j=1}^{r} \alpha_{1j} \circ Y_j \\
\vdots \\
\sum_{j=1}^{r} \alpha_{rj} \circ Y_j
\end{pmatrix}.
\]

The univariate operations \( \alpha \circ X \) and \( \beta \circ Y \) are independent if and only if the counting processes in their definitions are independent. Hence, the matrix-valued operator implies independence between the univariate operators. Properties of this operator can be found in Latour (1997).

Using this operator, Latour (1997) defined a multivariate generalized INAR process of order \( p \) (MGINAR(\( p \))) by assuming that

\[
Y_t = \sum_{j=1}^{p} A_j \circ Y_{t-j} + \epsilon_t,
\]

where \( Y_t \) and \( \epsilon_t \) are \( r \)-vectors and \( A_j, j = 1, \ldots, p \) are \( r \times r \) matrices and gave conditions for existence and stationarity. A more focused presentation of the model follows.

### 19.3.2 Multivariate INAR Model

Let \( Y \) and \( R \) be nonnegative integer-valued random \( r \)-vectors and let \( A \) be an \( r \times r \) matrix with elements \( \{\alpha_{ij}, i, j = 1, \ldots, r\} \). The MINAR(1) process can be defined as

\[
Y_t = A \circ Y_{t-1} + R_t = \begin{pmatrix}
\alpha_{11} & \alpha_{12} & \ldots & \alpha_{1r} \\
\alpha_{21} & \alpha_{22} & \ldots & \alpha_{2r} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{r1} & \alpha_{r2} & \ldots & \alpha_{rr}
\end{pmatrix} \circ \begin{pmatrix}
Y_{1,t-1} \\
\vdots \\
Y_{r,t-1}
\end{pmatrix} + \begin{pmatrix}
R_{1t} \\
\vdots \\
R_{rt}
\end{pmatrix}, \quad t = 1, 2, \ldots
\]

(19.3)

The vector of innovations \( R_t \) follows an \( r \)-variate discrete distribution, which characterizes the marginal distribution of the \( Y_t \) as well. More on this will follow. When \( A \) is diagonal, we will call the model diagonal MINAR. Clearly, this has less structure.
The nonnegative integer-valued random process \( \{Y_t\}_{t \in \mathbb{Z}} \) is the unique strictly stationary solution of (19.3), if the largest eigenvalue of the matrix \( A \) is less than 1 and \( E \|R_t\| < \infty \) (see also Franke and Rao, 1995; Latour, 1997).

To help the exposition consider the case with \( r = 2 \). The two series can be written as

\[
Y_{1t} = \alpha_{11} \circ Y_{1,t-1} + \alpha_{12} \circ Y_{2,t-1} + R_{1t}, \\
Y_{2t} = \alpha_{22} \circ Y_{2,t-1} + \alpha_{21} \circ Y_{1,t-1} + R_{2t}.
\]

This helps to understand the dynamics. The cross correlation between the two series comes from sharing common elements as well as from the joint distribution of \((R_{1t}, R_{2t})\).

If \( A \) is a diagonal matrix in this bivariate example, so that \( \alpha_{12} = \alpha_{21} = 0 \), then the two series are univariate INAR models but are still correlated due to the joint pmf of \((R_{1t}, R_{2t})\).

Taking expectations on both sides of (19.3), it is straightforward to obtain

\[
\mu = E(Y_t) = [I - A]^{-1} E(R_t).
\] (19.4)

The variance–covariance matrix \( \gamma(0) = E [(Y_t - \mu)(Y_t - \mu)'] \) satisfies a difference equation of the form

\[
\gamma(0) = A \gamma(0) A' + \text{diag}(B\mu) + \text{Var}(R_t).
\] (19.5)

The innovation series \( R_t \) consists of identically distributed sequences \( \{R_{it}\}_{i=1}^{r} \) and has mean \( E(R_t) = \lambda = (\lambda_1, \ldots, \lambda_r)' \) and variance

\[
\text{Var}(R_t) = \\
\begin{bmatrix}
\nu_1 \lambda_1 & \phi_{12} & \ldots & \phi_{1r} \\
\phi_{12} & \nu_2 \lambda_2 & \ldots & \phi_{2r} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{1r} & \phi_{2r} & \ldots & \nu_r \lambda_r
\end{bmatrix},
\]

where \( \nu_i > 0, i = 1, \ldots, r \). Depending on the value of the parameter \( \nu_i \), the assumptions of equidispersion (\( \nu_i = 1 \)), overdispersion (\( \nu_i > 1 \)), and underdispersion (\( \nu_i \in (0, 1) \)) can be obtained.

In the bivariate case, that is, when \( r = 2 \), it can be proved that the vector of expectations (19.5) has elements

\[
\mu_1 = \frac{(1 - \alpha_{22}) \lambda_1 + \alpha_{12} \lambda_2}{(1 - \alpha_{11})(1 - \alpha_{22}) - \alpha_{12} \alpha_{21}}, \\
\mu_2 = \frac{(1 - \alpha_{11}) \lambda_2 + \alpha_{21} \lambda_1}{(1 - \alpha_{11})(1 - \alpha_{22}) - \alpha_{12} \alpha_{21}},
\]

© 2016 by Taylor & Francis Group, LLC
while the elements of $\gamma(0)$ are

$$
\gamma_{11}(0) = \text{Var}(Y_{1t})
= \frac{1}{(1 - \alpha_{11}^2)} \left\{ \alpha_{12}^2 \text{Var}(Y_{2t}) + 2\alpha_{11}\alpha_{12} \text{Cov}(Y_{1t}, Y_{2t})
+ \alpha_{11}(1 - \alpha_{11})\mu_1 + \alpha_{12}(1 - \alpha_{12})\mu_2 + \nu_1\lambda_1 \right\},
$$

$$
\gamma_{22}(0) = \text{Var}(Y_{2t})
= \frac{1}{(1 - \alpha_{22}^2)} \left\{ \alpha_{21}^2 \text{Var}(Y_{1t}) + 2\alpha_{22}\alpha_{21} \text{Cov}(Y_{1t}, Y_{2t})
+ \alpha_{22}(1 - \alpha_{22})\mu_2 + \alpha_{21}(1 - \alpha_{21})\mu_1 + \nu_2\lambda_2 \right\},
$$

$$
\gamma_{12}(0) = \gamma_{21}(0) = \text{Cov}(Y_{1t}, Y_{2t})
= \frac{\alpha_{11}\alpha_{21} \text{Var}(Y_{1t}) + \alpha_{22}\alpha_{12} \text{Var}(Y_{2t}) + \phi}{1 - \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}},
$$

where $\phi$ is the covariance between the innovations.

Note that $\text{Cov}(Y_{it}, R_{jt}) = \text{Cov}(R_{it}, R_{jt})$, $i, j = 1, \ldots, r$, $i \neq j$ (Pedeli and Karlis, 2011). That is, the covariance between the current value of one process and the innovations of the other process at time $t$ is equal to the covariance of the innovations of the two series at the same time $t$.

Regarding the covariance function $\gamma(h) = E [(Y_{t+h} - \mu)(Y_t - \mu)']$ for $h > 0$, iterative calculations provide us with an expression of the form

$$
\gamma(h) = A \gamma(h - 1) = A^h \gamma(0), \quad h \geq 1, \quad (19.6)
$$

where $\gamma(0)$ is given by (19.5).

Applying the well-known Cayley–Hamilton theorem to (19.6), it is straightforward to show that the marginal processes will have an ARMA $(r, r - 1)$ correlation structure. Since $A$ is an $r \times r$ matrix, the Cayley–Hamilton theorem ensures that there exist constants $\xi_1, \ldots, \xi_r$, such that $A^r - \xi_1 A^{r-1} - \cdots - \xi_r I = 0$. Thus, $\gamma(h)$ satisfies

$$
\gamma(h) - \xi_1 \gamma(h - 1) - \cdots - \xi_r \gamma(h - r) = 0, \quad h \geq r. \quad (19.7)
$$

Equations (19.6) and (19.7) hold for every element in $\gamma(h)$, and hence, the autocorrelation function of $\{Y_{jt}\}, j = 1, \ldots, r$ satisfies

$$
\rho_{jj}(h) - \sum_{i=1}^r \xi_i \rho_{jj}(h - i) = 0, \quad h \geq r.
$$
Thus, each component has an ARMA($r,r-1$) correlation structure (see also McKenzie, 1988; Dewald et al., 1989). In the simplest case of a BINAR(1) model, the marginal processes have ARMA(2,1) correlations with $\xi_1 = \alpha_{11} + \alpha_{22}$ and $\xi_2 = \alpha_{12}\alpha_{21} - \alpha_{11}\alpha_{22}$. For the diagonal MINAR(p) case, the marginal process is the simple univariate INAR(p) process.

Al-Osh and Alzaid (1987) expressed the marginal distribution of the INAR(1) model in terms of the innovation sequence $\{R_t\}$, that is, $Y_t \overset{d}{=} \sum_{i=0}^{\infty} \alpha_i \circ R_{t-i}$. This result was easily extended to the case of a diagonal MINAR(1) process (Pedeli and Karlis, 2013c) where

$$Y_{jt} \overset{d}{=} \sum_{i=0}^{\infty} \alpha_{ij} \circ R_{j,t-i}.$$ 

For the general MINAR(1) process, the distribution of such a process can also be expressed in terms of the multivariate innovation sequence $R_t$ as

$$Y_t \overset{d}{=} \sum_{i=0}^{\infty} A^i \circ R_{t-i},$$ 

where $A^i = PD^iP^{-1}$. Here, $P$ is the matrix of the eigenvectors of $A$ and $D$ is the diagonal matrix of the eigenvalues of $A$. Since all the eigenvalues should be smaller than 1 in order for stationarity to hold, the matrix $D^i$ tends to a zero matrix as $i \to \infty$ and hence $A^i$ tends to zero as well.

The usefulness of such expressions is that they facilitate the derivation of the (joint) probability generating function (pgf) of the (multivariate) process, thus revealing its distribution. Assuming stationarity, the joint pgf $G_Y(s)$ satisfies the difference equation

$$G_Y(s) = G_Y(A^Ts)G_R(s).$$

More details can be found in Pedeli and Karlis (2013c).

Extensions of the model mentioned earlier are possible. One can add covariates to the mean of the innovations using a log link function. This allows us to fit the effect of some other covariates to the observed multivariate time series, see Pedeli and Karlis (2013b) for such an application. Also, extensions to higher order are straightforward but lead to rather complicated models.

19.3.3 Estimation

The least squares approach for estimation was discussed in Latour (1997). However, based on parametric assumptions for the innovations, other estimation methods are available. Parametric models also offer more flexibility for predictions.

For the estimation of the BINAR(1) model, the method of conditional maximum likelihood can be used. The conditional density of the BINAR(1) model can be constructed as the convolution of
\begin{align*}
f_1(k) &= \sum_{j_1=0}^{k} \binom{y_{1,t-1}}{j_1} \binom{y_{2,t-1}}{k-j_1} \alpha_{11}^{j_1} (1 - \alpha_{11})^{y_{1,t-1} - j_1} \alpha_{12}^{k-j_1} (1 - \alpha_{12})^{y_{2,t-1} - k+j_1}, \\
f_2(s) &= \sum_{j_2=0}^{s} \binom{y_{2,t-1}}{j_2} \binom{y_{1,t-1}}{s-j_2} \alpha_{22}^{j_2} (1 - \alpha_{22})^{y_{2,t-1} - j_2} \alpha_{21}^{s-j_2} (1 - \alpha_{21})^{y_{1,t-1} - s+j_2},
\end{align*}

and a bivariate distribution of the form \( f_3(r_1, r_2) = P(R_{1t} = r_1, R_{2t} = r_2) \). The functions \( f_1(\cdot) \) and \( f_2(\cdot) \) are the pmfs of a convolution of two binomial variates. Thus, the conditional density takes the form

\[ f(y_t | y_{t-1}, \theta) = \sum_{k=0}^{s_1} \sum_{s=0}^{s_2} f_1(k) f_2(s) f_3(y_{1t} - k, y_{2t} - s), \]

where \( s_1 = \min(y_{1t}, y_{1,t-1}) \) and \( s_2 = \min(y_{2t}, y_{2,t-1}) \). Maximum likelihood estimates of the vector of unknown parameters \( \theta \) can be obtained by maximization of the conditional likelihood function

\[ L(\theta | y) = \prod_{t=1}^{T} f(y_t | y_{t-1}, \theta) \quad (19.8) \]

for some initial value \( y_0 \). The asymptotic normality of the conditional maximum likelihood estimate \( \hat{\theta} \) has been shown in Franke and Rao (1995) after imposing a set of regularity conditions and applying the results of Billingsley (1961) for the estimation of Markov processes.

Numerical maximization of (19.8) is straightforward with standard statistical packages. The binomial convolution implies finite summation and hence it is feasible. Note also that since the pgf of a binomial distribution is a polynomial, one can derive the pmf of the convolution easily via polynomial multiplication using packages in R. Depending on the choice for the innovation distribution, the conditional maximum likelihood (CML) approach can be applied. In Pedeli and Karlis (2013c), a bivariate Poisson and a bivariate negative binomial distribution were used. For the parametric models prediction was discussed. An interesting result is that for the bivariate Poisson innovations the univariate series have a Hermite marginal distribution. In Karlis and Pedeli (2013), a copula-based bivariate innovation distribution was used allowing negative cross-correlation.

When moving to the multivariate case things become more demanding. First of all, a multivariate discrete distribution is needed for the innovations. As discussed in Section 19.2, such models can be complicated. In Pedeli and Karlis (2013a), a multivariate Poisson distribution is assumed with a diagonal matrix \( A \). Even in this case, the pmf of the multivariate Poisson distribution is demanding since multiple summation is needed. The conditional likelihood can be derived as in the bivariate case but now this is a convolution of several binomials and a multivariate discrete distribution. Alternatively, a composite likelihood approach can be used. Composite likelihood methods are based on the idea of constructing lower-dimensional score functions that still contain enough information about the structure considered but they are computationally more tractable (Varin, 2008). See also Davis and Yau (2011) for asymptotic properties of composite likelihood methods applied to linear time series models.
Application of composite likelihood approach implies the usage, for example, of bivariate marginal log-likelihood functions over all pairs instead of the usage of the multivariate likelihood. As an illustration, consider a trivariate probability function \( P(x, y, z; \theta) \), where \( \theta \) is a vector of parameters to estimate. The log-likelihood to be maximized is of the form
\[
\ell(\theta) = \sum_{i=1}^{n} \log P(x_i, y_i, z_i; \theta)
\]
while the composite log-likelihood is
\[
\ell_c(\theta) = \sum_{i=1}^{n} \left[ \log P(x_i, y_i; \theta) + \log P(x_i, z_i; \theta) + \log P(y_i, z_i; \theta) \right],
\]
that is, we replace the trivariate distribution by the product of the bivariate ones. The price to be paid is some loss of efficiency which can be very large for some models. On the other hand, the optimization problem is usually easier. Simulations for the diagonal trivariate MINAR(1) model have shown small efficiency loss (Pedeli and Karlis, 2013a). Therefore, the composite likelihood method makes feasible the application of some multivariate models for time series and this is worth further exploration. Alternatively, one may employ an expectation-maximization (EM) algorithm making use of the latent structure imposed by the convolution (Pedeli and Karlis, 2013a).

Finally, Bayesian estimation of the BINAR model with bivariate Poisson innovations is described in Sofronas (2012).

### 19.3.4 Other Models in This Category

Ristic et al. (2012) developed a simple bivariate integer-valued time series model with positively correlated geometric marginals based on the negative binomial thinning mechanism. Bulla et al. (2011) described a model based on another operator called the signed binomial operator allowing fitting integer-valued data in \( \mathbb{Z} \). Recently, Scotto et al. (2014) derived a model for correlated binomial data and Nastic et al. (2014) discussed a model with Poisson marginals with same means.

Quoreshi (2006, 2008) described properties of a bivariate moving-average model (BINMA). The BINMA(\( q_1, q_2 \)) model takes the form
\[
y_{1t} = u_{1t} + a_{11} \circ u_{1,t-1} + \cdots + a_{1q_1} u_{1,t-q_1} \\
y_{2t} = u_{2t} + a_{21} \circ u_{2,t-1} + \cdots + a_{2q_2} u_{2,t-q_2},
\]
where \( u_{ij} \) are innovation terms following some positive discrete distribution. Estimation using conditional least squares, feasible least squares, and generalized method of moments is described. Cross-correlation between the series is implied by assuming dependent innovations terms. Extensions to the multivariate case are described in Quoreshi (2008). Similar to the MINAR model, the multivariate vector integer MA model can allow for mixing between the series. In both the bivariate and the multivariate models, no parametric assumptions for the innovations were given. A BINMA model has been studied in Brännäs and Nordström (2000).
19.4 Parameter-Driven Models

Parameter-driven models have also been considered for count time series. The main idea is that the serial correlation imposed is due to some correlated latent processes on the parameter space. They offer useful properties since the serial correlation of the latent process drives the serial correlation properties for the count model. On the other hand, estimation is usually much harder. We describe such multivariate models in the following.

19.4.1 Latent Factor Model

Jung et al. (2011) presented a factor model which in fact belongs to the family of parameter-driven models. The model was applied to the number of trades of five stocks belonging to two different sectors within a 5 min interval for a period of 61 days. The model assumed that the number of trades $y_{it}$ for the $i$th stock at time $t$ follows a Poisson distribution with mean $\theta_{it}$, while

$$
\log \theta_{it} = \mu_i + \gamma \lambda_t + \delta_i \tau_{si,t} + \phi_i \omega_{it},
$$

where $\mu_i$ is a mean specific to the $i$th stock, which perhaps may relate to some covariates specific to the stock as well, $\lambda_t$ is a latent common market factor, $\tau_t = (\tau_{1t}, \tau_{st})'$ is a latent vector of industry-specific factors and $\omega_t = (\omega_{1t}, \ldots, \omega_{Jt})$ are latent stock-specific factors.

The model assumes an AR(1) specification for the latent factors, namely,

$$
\lambda_t \mid \lambda_{t-1} \sim N\left(\kappa_{\lambda} + \nu_{\lambda} \lambda_{t-1}, \sigma_{\lambda}^2\right)
$$

$$
\tau_{st} \mid \tau_{s,t-1} \sim N\left(\kappa_{\tau_s} + \nu_{\tau_s} \tau_{s,t-1}, \sigma_{\tau_s}^2\right)
$$

$$
\omega_{it} \mid \omega_{i,t-1} \sim N\left(\kappa_{\omega_i} + \nu_{\omega_i} \omega_{i,t-1}, \sigma_{\omega_i}^2\right).
$$

Clearly, cross-correlation enters the model by assuming common factors, while serial correlation from the latent AR processes. The likelihood is complicated, and the authors developed an efficient importance sampling algorithm to evaluate the log-likelihood and apply the simulated likelihood approach.

19.4.2 State Space Model

Jorgensen et al. (1999) proposed a multivariate Poisson state space model with a common factor that can be analyzed by a standard Kalman filter. The model assumes that the multiple counts $y_{it}$, $i = 1, \ldots, d$ at time $t$ follow conditionally independent Poisson distributions, namely,

$$
Y_{it} \sim \text{Poisson}(\alpha_{it} \theta_t)
$$
with $\alpha_i = \exp(x'_i\alpha_i)$, where $x_i$ is a vector of time-varying covariates including a constant and further

$$\theta_i | \theta_{i-1} \sim \text{Gamma} \left( b_i \theta_{i-1}, \frac{\sigma^2}{\theta_{i-1}} \right),$$

where $\text{Gamma}(a, b^2)$ is a gamma distribution with mean $a$ and coefficient of variation $b$. Kalman filtering is used for the latent process.

Finally, Lee et al. (2005) proposed a model starting from a bivariate zero-inflated Poisson regression with covariates, adding random effects that are correlated in time.

### 19.5 Observation-Driven Models

In observation-driven models, serial correlation comes from the fact that current observations relate directly to the previous ones. The Poisson autoregression model defined in Fokianos et al. (2009) constitutes an important member of this class for univariate series. The model has a feedback mechanism and is defined as

$$Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \lambda_t = d + a \lambda_{t-1} + b Y_{t-1},$$

for $t \geq 1$, where the parameters $d, a, b$ are assumed to be positive. In addition, assume that $Y_0$ and $\lambda_0$ are fixed and $\mathcal{F}_{t-1}$ is the information up to time $t - 1$. The model is called INGARCH, but the name perhaps is very ambitious. Properties of the model can be seen in Fokianos et al. (2009). The model has found a lot of work after this (Trostheim, 2012; Davis and Liu, 2015). The Poisson INGARCH is incapable of modeling negative serial dependence in the observations which is, however, possible by the self-excited threshold Poisson autoregression model (see Wang et al., 2014). Extensions to higher-order INGARCH($p,q$) (see Weiβ, 2009), nonlinear relationships, and other distributional assumptions have been also proposed in the literature. This type of model offers a much richer autocorrelation structure than models based on thinning, like long memory properties, for example. Their estimation is more computational demanding, the same is true for deriving their properties.

Extension to higher dimensions has also been proposed. Liu (2012) proposed a bivariate Poisson integer-valued GARCH (BINGARCH) model. This model is capable of modeling the time dependence between two time series of counts. Consider two time series $Y_{1t}$ and $Y_{2t}$. We assume that

$$Y_t = (Y_{1t}, Y_{2t}) | \mathcal{F}_{t-1} \sim \text{BP}_2(\lambda_{1t}, \lambda_{2t}, \phi),$$

where $\text{BP}_2(\lambda_{1t}, \lambda_{2t}, \phi)$ denotes a bivariate Poisson distribution with marginal means $\lambda_{1t}$ and $\lambda_{2t}$, respectively, and covariance equal to $\phi$. This is a reparameterized version of the distribution in (19.1). Furthermore, we assume for the general BIV. INGARCH($m,q$) model that

$$\lambda_t = \delta + \sum_{i=1}^m A_i \lambda_{t-i} + \sum_{j=1}^q B_j Y_{t-j},$$
\( \lambda_t = (\lambda_{1t}, \lambda_{2t})', \delta > 0 \) is a 2-vector and \( A_i \) and \( B_j \) are \( 2 \times 2 \) matrices with nonnegative entries. Conditions to ensure the positivity of \( \lambda \) are given.

For example, the BIV.INGARCH(1,1) model takes the form

\[
\lambda_t = \delta + A \lambda_{t-1} + B Y_{t-1}
\]

or equivalently

\[
\begin{pmatrix} \lambda_{1t} \\ \lambda_{2t} \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{pmatrix} \lambda_{1,t-1} \\ \lambda_{2,t-1} \end{pmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix}.
\]

Stability properties for this model are not simple. Liu (2012) used the iterated random functions approach that allows us to derive the stability properties under a contracting constraint on the coefficient matrices. Inference procedures are also presented and applied to real data in the area of traffic accident analysis.

Heinen and Rengifo (2007) developed a model called the Autoregressive Conditional double Poisson model for a \( d \)-dimensional vector of counts \( y_t \). The model relates to the one given earlier. Their model was based on the double Poisson distribution defined in Efron (1986):

\[
Y_{it} | F_{t-1} \sim DP(\lambda_{it}, \phi_i), \quad i = 1, \ldots, d,
\]

where \( \lambda_{it} \) is the mean of the double Poisson distribution of the \( i \)th count at time \( t \) and \( \phi_i \) is an overdispersion parameter. For \( \phi_i = 1 \), we get the Poisson distribution. Then the mean vector \( \lambda_t = (\lambda_{1t}, \ldots, \lambda_{dt}) \) is defined as a VARMA process

\[
\lambda_t = \omega + A \mu_{t-1} + B Y_{t-1},
\]

where \( A \) and \( B \) are appropriate matrices. Stationarity is ensured as long as the eigenvalues of \( (I - A - B) \) lie within the unit circle. The VARMA order specification can be modified. The cross-correlation between the counts is imposed via a Gaussian copula. The model uses a trick to avoid problems when working with discrete-valued data and copulas, by using a continued extension argument, that is, by adding some noise to the counts to make them continuous and working with the continuous versions. The latter adjustment may create some noise around the model. A recent paper (Nikoloulopoulos, 2013a) discusses the problems with such a method.

Bien et al. (2011) used copulas to create a multivariate time series model defined on \( \mathbb{Z} \) and not only on positive integers. They modeled a bivariate time series of bid and ask quote changes sampled at a high frequency. The marginal models used were assumed to follow a dynamic integer count hurdle (ICH) process, tied together with a copula, which was constructed by properly taking into account the discreteness of the data.

Finally, Held et al. (2005) described a model with multivariate counts where at time \( t \) counts of the other series at time \( t - 1 \) enter as covariates for the mean of each series. Also, Brandt and Sandler (2012) used the model of Chib and Winkelmann (2001) and made it dynamic by adding autoregressive terms in the mean of the mixing distribution.
19.6 More Models

The earlier mentioned list of models does not limit other families of models to be derived. Such an example is the creation of hidden markov models (HMMs) for multivariate count time series. In the univariate case, Poisson HMMs have been used to model integer-valued time series data. Extensions to the multivariate case are possible but multivariate discrete distributions are needed to model the state distribution. In Orfanogiannaki and Karlis (2013), multivariate Poisson distributions based on copulas have been considered for this purpose.

Another approach to create multivariate time series models for counts could be through the discretization of standard continuous models. Consider, for example, the vector autoregressive model based on a bivariate normal distribution. Discretizing the output can lead to the desired time series. However, such a discretization is not unique and perhaps problems may occur while estimating the parameters, as multivariate integrals need to be calculated.

Joe (1996) described an approach to create time series models based on additively closed families. Working with bivariate distributions with this property, as, for example, the bivariate Poisson distribution, one can derive such a time series model. Such models share common elements with models based on thinning operations; see Joe (1996) for a methodology to create an appropriate operator.

19.7 Discussion

In this chapter, we have pulled together the existing literature on multivariate integer-valued time series modeling. Table 19.1 summarizes the models. An obstacle for such models is the lack of, or at least the lack of familiarity with, multivariate discrete distributions which are basic tools for their construction. Given greater availability of such basic tools, we expect that more models will become available in the near future. Also, ideas for tackling estimation problems like the composite likelihood approach can help a lot to this direction.

Such models can have also some other interesting potential. For example, taking the difference of the two time series from a bivariate model, we end up with a time series

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models based on thinning</td>
<td>Franke and Rao (1995); Latour (1997); Pedeli and Karlis (2011); Pedeli and Karlis (2013a,c); Pedeli and Karlis (2013b); Karlis and Pedeli (2013); Boudreault and Charpentier (2011); Quoreshi (2006, 2008); Ristic et al. (2012); Brännäs and Nordström (2000); Bulla et al. (2011)</td>
</tr>
<tr>
<td>Observation-driven models</td>
<td>Liu (2012); Heinen and Rengifo (2007); Bien et al. (2011); Held et al. (2005); Brandt and Sandler (2012)</td>
</tr>
<tr>
<td>Parameter-driven models</td>
<td>Jung et al. (2011); Jorgensen et al. (1999); Lee et al. (2005)</td>
</tr>
</tbody>
</table>
defined on $\mathbb{Z}$. Models for time series on $\mathbb{Z}$ are becoming popular in several disciplines and some of them can be derived by higher-dimension models. Multivariate time series in $\mathbb{Z}^d$ would also be of interest. For example, in finance one needs to model the number of ticks that a stock is going up or down during consecutive time points, and this can create a large number of time series, when managing a portfolio.

**Acknowledgment**

The author would like to thank Dr. Pedeli for helpful comments during the preparation of this chapter.

**References**


