International Journal of Statistics and Probability; Vol. 11, No. 6; November 2022
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Review of Copula for Bivariate Distributions of Zero-Inflated Count Time Series Data

Dimuthu Fernando[1], Mohammed Alqawba[2], Manar Samad[3], Norou Diawara[1]

[1.] Department of Mathematics & Statistics, College of Sciences, Old Dominion University, Norfolk VA, USA

[2.] Department of Mathematics, College of Science and Arts, Qassim University, Ar Rass 51452, Saudi Arabia

[3.]Dept. of Computer Science, Tennessee State University, Nashville, TN, USA

Correspondence: Norou Diawara, Department of Mathematics & Statistics, College of Sciences, Old Dominion University, Norfolk VA, USA

Received: August 29, 2022 Accepted: October 11, 2022 Online Published: October 26, 2022

doi:10.5539/ijsp.v11n6p28 URL: https://doi.org/10.5539/ijsp.v11n6p28

Abstract

The class of bivariate integer-valued time series models, described via copula theory, is gaining popularity in the literature because of applications in health sciences, engineering, financial management and more. Each time series follows a Markov chain with the serial dependence captured using copula-based distribution functions from the Poisson and the zero-inflated Poisson margins. The copula theory is again used to capture the dependence between the two series.

However, the efficiency and adaptability of the copula are being challenged because of the discrete nature of data and also in the case of zero-inflation of count time series. Likelihood-based inference is used to estimate the model parameters for simulated and real data with the bivariate integral of copula functions. While such copula functions offer great flexibility in capturing dependence, there remain challenges related to identifying the best copula type for a given application. This paper presents a survey of the literature on bivariate copula for discrete data with an emphasis on the zero-inflated nature of the modelling. We demonstrate additional experiments on to confirm that the copula has potential as greater research area.

Keywords: count time series, copula, Zero-Inflated, count data, Poisson distribution

Subject Classification: 62H05, 62H10

1. Introduction

In the study of multivariate distributions, copula functions are gaining popularity in recent years. They are attractive as they can handle internal and mutual dependences among variables. The copula was first introduced in the Sklar (1954) paper, a paper that Frechet helped publish. Hoffding (1940) is also credited for almost innovating the concept of copula . Many problems in practical situations are modeled under related distributions using copula functions, in contrast to classical multivariate (Gaussian) distributions for count data. As such, the literature shows a growing interest in the investigation of dependence for sequences of counts in time series cases. The simplest of such sequences are bivariate count time series data. Copula functions have gained popularity in building such bivariate and multivariate distributions as the desire to understand the structure in massive time series count data is becoming more common. For diseases and rare events, observed counts over time appear in a high frequency of zeros (zero inflation), which is discussed in MÖller et al. (2020) and Young et al. (2020).

Sklar (1959) introduced a method to build in the bivariate and multivariate distributions for two random variables. The idea of joint distribution, especially in the bivariate case can be traced back to Frechet (1951, 1956, 1958). Morgensetrn (1956), Plackett (1965), Farlie (1969) and many other authors could be included in this systematic approach of constructing bivariate distributions with specific marginals and different dependence measures. See examples such as Gumbel (1958) or Johnson and Tenenbein (1981). In that same line of thought, Cook and Johnson (1981) asked two questions that are still of relevance. The questions are: 1) “Is there a distribution that appears to be the most promising candidate for non-normal types of data?” 2) “Is the resulting distribution or model fit significantly better than that obtained from the multivariate normal distribution?”

Finding a unique copula for a joint distribution requires one to know the form of the joint distribution. When using copula, one can separately model the marginal distributions and the dependence structure, which makes the copula approach unique. Choosing the appropriate copula for a particular scenario means finding the one that best captures the dependence in data. Many variants of copulas have been proposed in the literature where each of these is suitable for different dependence structures. For example, Gaussian copula is flexible, and it allows for equally positive and negative dependence. The Clayton copula cannot account for negative dependence, and it exhibits strong left tail dependence. Similar to Gaussian copula, Frank copula allows for both positive and negative dependence between the marginals.

Copulas offer a flexible framework to combine distributions. It is unique if marginal densities are continuous. However, if some of the marginal distributions are discrete, the unicity cannot be obtained automatically.

Many copula functions have been identified, from the extreme of independent variables (the so-called independent copula or the product) to the max or min copula. The dependence is then captured by a selection of parameters and criteria associated with the range and properties of model parameters.

Moreover, high dimensional copulas have been introduced via bivariate copulas, under different decompositions and structures. These structures are known as the canonical vine (C-vine) or drawable vine (D-vine). References to C and D vines can be found in Bedford and Cooke (2002), Joe et al. (2010), and Aas et al. (2009). Gräler (2014) proposed the convex combination of bivariate copula densities incorporating the distance [between what?] as a parameter in the spatial setting. The application of copula functions can be found in finances (Czado et al, 2012), hydrology (Yu et al., 2020), transportation (Irannezhad et al., 2017), health care (Shi and Zhang, 2015), and more. The Farlie-Gumbel-Morgenstern (FGM) family of copula can be used to establish relationship between predictors (Durante and Sempi, 2016)).

Within the count time series, if we look at the binary data, there is a growing interest in the description of multivariate distributions under pair copulas (Lin and Chaganty, 2021). Panagiotelis et al. (2012) presented pair copula constructions for discrete multivariate data. Their algorithm is explained as a product of bivariate pair copula, demonstrating the great potential of vine copula approaches. They stated that the model selection for C or D vine remains an important open problem, with a particular emphasis on the conditional independence identification (Czado, 2019, Deng and Chaganty, 2021,). From there, the idea of using the D vine for modeling counts with excess zeros and temporal dependence is presented in Sefidi et al. (2020). Perrone and Durante (2021) highlighted the link between the extreme discrete copula and mathematical concept of convex polytope, which is an idea spinning from the class of bivariate distributions (Rao and Subramanyam (1990).

There are numerous problems and interesting challenges related to time series of counts. Davis et al. (2016, 2021) presented extensive literature and many examples of count time series. Fokianos (2021) and Armillota and Fokianos (2021) presented a Poison network autoregression for counts. In the statistical process control, Fatahi et al. (2012) proposed the monitoring of rare events under the copula based bivariate zero-inflated Poisson. van Den Heuvel et al. (2020) proposed corrections to such results adding the negative correlation option.

With these studies and observations in mind, this paper presents reviews and updates related to the copula for bivariate distributions of zero-inflated count time series and highlights research directions. Motivated by multivariate datasets acquired using correlation structures, our goal is to review the bivariate count and zero-inflated count time series for inference and application purposes under copula modeling. We give some insights into the bivariate count copula and its recent developments. We organize our discussion as follows. In Section 2, copulas for discrete count and zero-inflation of discrete count time series data are described. The use of univariate and bivariate copula for discrete data is discussed in Section 3. Extensions of discrete bivariate copulas are described in Section 4. We conclude this paper with an extended discussion on future work.

2. Copula for Zero-inflated of Discrete and Count Time Series Data

This section introduces the general form for multivariate copula, and its Gaussian representation. We also give an explicit definition of the zero inflated counts time series data.

2.1 Simple Gaussian Copula Example

Masarotto and Varin (2012) introduced a Gaussian copula model which can be used to model time series data in the presence of covariates. The corresponding regression model can be written as follows.

$$\mathrm{Y_t} = \mathrm{g}(X_t, \epsilon_\mathrm{t}\ \theta), \mathrm{for\ t} = 1, .., \mathrm{n},$$

where $g(.)$ is a function of the covariates X_t and ϵ_t, which capture the serial dependence. The parameter $\boldsymbol{\theta}$ is a vector of marginal regression coefficients. The joint distribution function of the time series $\{Y_t\}\ for\ t = 1, \dots, n$ can be constructed using the Gaussian copula as follows.

$$F(y_1, y_2, \dots, y_n) = P(Y_1 \le y_1, Y_2 \le y_2, \dots, Y_n \le y_n) = \Phi_{R(\rho)}(\Phi^{-1}(F_1(y_1)), \Phi^{-1}(F_2(y_2)) \dots, \Phi^{-1}(F_n(y_n))) \quad (1)$$

Here, Φ^{-1} is the inverse CDF of standard normal distribution, and $\Phi_{R(\rho)}$ is the joint CDF of a multivariate normal distribution with a mean vector of zeros and covariance matrix R.

2.2 Review of Copula for Discrete Data

Copula distributions are becoming increasingly popular in many areas of statistical data sciences. For example, in engineering, copula distributions are used to model the shear fore for cantilever beams and for beams with multiple point loads (Zhang and Lam, 2016). In pharmaceutical quality control, two correlated characteristics sample data are presented in Fatahi et al. (2012). The authors describe the bivariate Poisson distribution with the evidence of zero-inflation. Sukparungsee et al. (2021) developed a bivariate copula for control chart effectiveness. They show the bivariate copula distribution on Hotelling's T^2 over the multivariate cumulative sum for positive, negative, weak, moderate, and strong correlations when the assumption of multivariate normality is violated. Van den Heuvel et al. (2020) extended the idea from Fatiha et al. (2012) and included negative correlation case, and an upper control limit on the sum of bivariate random variables. Copulas are elegantly captured in the Genest and MacKay (1986), Genest (1987) and also in Han and De Oliveira (2016 and 2020), among others. In the financial sector, a recent work by Nikoloulopoulos and Moffatt (2019) reminds us of the need to study dependence structures. There are also more general ambitions for the bivariate copula from a bigger perspective than we expect to show the aggregated effects in many other areas.

The list of copula functions is very large. The work of Größer and Okhrin (2021) presents a summary of bivariate copula followed by the construction of multivariate copula using pair copula decompositions. They provide examples for each copula family and provide an overview of how copula theory can be used in various fields of data science.

Yang et al. (2014) proposed the Ali-Mikhael-Haq (AMH) copula-based function to investigate the joint risk probabilities of rainstorms, wind speeds, and storm surges. The proposed model was developed to assess the impact based on marginal distributions of maximum daily rainfall and extreme gust velocity. Alqawba et al. (2021) constructed a class of bivariate integer-valued time series models using copula theory. Applying either the bivariate Gaussian copula or the bivariate t copula functions, they jointly modeled two copula-based Markov time series models. They applied their method on bivariate count time series data, where the marginals follow either a Poisson or zero-inflated Poisson distribution.

Safari et al. (2020) proposed a bivariate copula regression model to analyze cervical cancer data. They applied a bivariate copula to model and estimate joint distribution parameters. Nikoloulopoulos and Moffatt (2019) used bivariate copulas to jointly model bivariate ordinal time-series responses with covariates for risks assessment of married couples. They proposed a copula-based Markov modelling of ordinal time-series responses and used another copula to couple their conditional (on the past) distributions at each time point. Copula families such as the Bivariate normal (BVN), Frank, Gumbel and bivariate t-copula were used to model the univariate time series as well as to couple them together.

The work of Nikoloulopoulos & Karlis (2010) presents a regression copula-based model where covariates are used not only for the marginal but also for the copula parameters. They measured the effect of covariates on dependence structure by building a fully parametric copula-based model while considering six one-parameter copula families, namely Frank, Galambos, Gumbel, Mardia–Takahasi (M-T), and normal to build the dependence structure.

Karlis & Pedeli (2013) presented a bivariate integer-valued autoregressive process (BINAR(1)) in which the cross-correlation was modeled using a copula to accommodate both positive and negative correlation. They presented an application of the Frank and Gaussian copula to model dependence, and marginal time series were modeled using Poisson and negative binomial INAR(1) distributions.

Ma et al. (2020) proposed a copula approach utilizing a Gaussian copula with random effects to model correlated bivariate count data regression.

2.3 The Zero-Inflated Discrete Data

Zero inflation models can be found in many studies from Lambert (1992) to Hall (2000) and recently in Rigby et al. (2019). The zero-inflated count regression models are described as follows.

- Zero-Inflated Poisson (ZIP) Distribution (Lambert, 1992):

$$F_{Y_t}(m) = \omega_t + (1-\omega_t)e^{-\lambda_t}\sum_{y_t=0}^{m}\frac{\lambda_t^{y_t}}{y_t!}. \quad (2)$$

- Zero-Inflated Negative Binomial (ZINB) Distribution (Ridout et al, 2001):

$$F_{Y_t}(m) = \omega_t + \frac{(1-\omega_t)}{\Gamma(\kappa_t)}\left(\frac{\kappa_t}{\kappa_t+\lambda_t}\right)^{\kappa_t} \sum_{y_t=0}^{m} \frac{\Gamma(\kappa_t+y_t)}{y_t!} \left(\frac{\lambda_t}{\kappa_t+\lambda_t}\right)^{y_t}.$$

- Zero-Inflated Conway-Maxwell-Poisson (ZICMP) Distribution (Sellers and Raim, 2016):

$$F_{Y_t}(m) = \omega_t + \frac{(1-\omega_t)}{Z(\lambda_t+\kappa_t)} \sum_{y_t=0}^{m} \frac{\lambda_t^{y_t}}{(y_t!)^{\kappa_t}},$$

where $\lambda_t = \exp(\boldsymbol{X}'_t\beta)$, $\omega_t = \frac{\exp(\boldsymbol{Z}'_t\gamma)}{1+\exp(\boldsymbol{Z}'_t\gamma)}$, and $\kappa_t = \exp(\boldsymbol{W}'_t\alpha)$

are the associated covariate vectors affecting the intensity parameter λ_t, the zero-inflation parameter ω_t and the dispersion parameter κ_t , respectively.

The term $\sum_{y_t=0}^{m} \frac{\lambda_t^{y_t}}{(y_t!)^{\kappa_t}}$ is the normalizing function of the CMP.

Different variants of similar regression models have been proposed in the literature. A noteworthy use of copula for zero-inflated data is studied in Shamma et al. (2020), where the inflation is built from a geometric count time series in an integer-valued autoregressive (INAR) process.

3. Univariate and Bivariate Copula Models for Count Time Series Data

3.1 Univariate Copula-Based Model for Count Time Series Data

First order Markov model

Alqawba, & Diawara (2021) introduced a class of Markov zero inflated count time series model where the joint distribution function of the consecutive observations is constructed through copula functions. Suppose $\{Y_t\}$ zero-inflated count time series first order Markov chains the multivariate joint density distribution of $Y_1, Y_2, \ldots, Y_n$ can be constructed as below.

$$Pr(Y_1 = y_1, \ldots, Y_n = y_n) = Pr(Y_1 = y_1) \prod_{t=2}^{n} \Pr(Y_t = y_t | Y_{t-1} = y_{t-1})$$

Using the copula theory, the joint distribution function of Y_t, Y_{t-1} can be written as below.

$F_{12}(y_t \; y_{t-1}) = C(F_t(yt), F_{t-1}(y_{t-1}); \boldsymbol{\delta})$ *where* $\boldsymbol{\delta}$ *is bivariate copula parameter vector.*

Hence, we can calculate the transition probability as below.

$$\Pr(Y_t = y_t | Y_{t-1} = y_{t-1}) = \frac{\Pr(Y_t = y_t, Y_{t-1} = y_{t-1})}{f_{t-1}(y_{t-1})}$$

Where

$$P\,r(Y_1 = y_1, Y_{t-1} = y_{t-1}) = F_{12}(y_t, y_{t-1}) - F_{12}(y_t - 1, y_{t-1}) - F_{12}(y_t, y_{t-1} - 1) + F_{12}(y_t - 1, y_{t-1} - 1)$$

Likelihood and parameter estimation under first order Markov model

The likelihood function of the first order Markov model is given by

$$L(\vartheta, y) = \Pr(Y_1 = y_1; \theta) \prod_{t=2}^{n} \Pr(Y_t = y_t | Y_{t-1} = y_{t-1}; \vartheta) \quad (3)$$

The log likelihood function ($l(\vartheta; y)$) is given by

$$l(\vartheta; y) = \log \Pr(Y_1 = y_1; \theta) + \sum_{t=2}^{n} log \Pr(Y_t = y_t | Y_{t-1} = y_{t-1}; \vartheta)$$

Where θ *and* δ are the parameter vectors of the marginals and the dependence structure, respectively. For the Gaussian copula family, the likelihood function involves a bivariate integral of the normal probability in $C(.;\delta)$ which means that the function is not in a closed form and we need approximations for the rectangle probabilities.

The simulation study was conducted using the **R software** by the **'optim'** function in the **"stats" package.** We simulate first order stationary Markov processes with joint distribution of consecutive observations following the bivariate Gaussian copula. The marginal distributions are chosen to be the Poisson and ZIP distributions. We present the simulation results for a first order Markov model with Poisson marginals. The parameter λ represents the mean of a marginal Poisson, ω is the measure of zero inflation, and δ is the serial dependence associated with time series data.

We found that the estimate of these parameters is fairly stable where the precision increases with increasing sample size. Table 1 and Table 2 show the estimates of copula parameters for positive and negative autocorrelations, respectively. The estimates are described by standard measures of variation, including standard deviation, mean square error and mean absolute error.

Univariate ZI count time series models

For positive serial dependence with $\boldsymbol{\lambda}$**=3,** $\boldsymbol{\omega}$**=0.3,** $\boldsymbol{\delta}$ **=0.6**

Table 1. Parameter estimates for the univariate ZI Poisson model with positive autocorrelation

Sample Size	Parameters	Estimate	SE	MSE	MAE
100	$\boldsymbol{\lambda(3)}$	2.990	0.347	0.1200	0.282
	$\boldsymbol{\omega(0.3)}$	0.288	0.083	0.0070	0.006
	$\boldsymbol{\delta(0.6)}$	0.577	0.091	0.0080	0.073
300	$\boldsymbol{\lambda(3)}$	3.013	0.192	0.037	0.152
	$\boldsymbol{\omega(0.3)}$	0.293	0.046	0.002	0.037
	$\boldsymbol{\delta(0.6)}$	0.596	0.046	1.433	1.196
500	$\boldsymbol{\lambda(3)}$	3.006	0.154	0.024	0.120
	$\boldsymbol{\omega(0.3)}$	0.295	0.035	0.001	0.028
	$\boldsymbol{\delta(0.6)}$	0.596	0.037	0.001	0.028

Source: Fernando, D., Alqawba, M., Fernando, D., Diawara, N.& Samad, M. (2022).

For negative serial dependence with $\boldsymbol{\lambda}$**=3,** $\boldsymbol{\omega}$**=0.3,** $\boldsymbol{\delta}$ **=-0.6**

Table 2. Parameter estimates for the univariate ZI Poisson model with negative autocorrelation

Sample Size	Parameters	Estimate	SE	MSE	MAE
100	$\boldsymbol{\lambda(3)}$	3.045	0.280	0.080	0.234
	$\boldsymbol{\omega(0.3)}$	0.299	0.046	0.002	0.036
	$\boldsymbol{\delta(-0.6)}$	-0.618	0.087	0.0070	0.072
300	$\boldsymbol{\lambda(3)}$	3.019	0.152	0.023	0.119
	$\boldsymbol{\omega(0.3)}$	0.298	0.030	0.0007	0.002
	$\boldsymbol{\delta(-0.6)}$	-0.605	0.050	0.003	0.040
500	$\boldsymbol{\lambda(3)}$	3.014	0.112	0.0127	0.009
	$\boldsymbol{\omega(0.3)}$	0.299	0.019	0.0004	0.015
	$\boldsymbol{\delta(-0.6)}$	-0.603	0.040	0.002	0.031

Source: Fernando, D., Alqawba, M., Fernando, D., Diawara, N.& Samad, M. (2022).

Applications

Alqawba & Diawara (2021) applied the proposed model to analyze monthly count of strong sandstorms recorded by the AQI airport station in Eastern Province, Saudi Arabia. The data set consists of 348 monthly counts of strong sandstorms, starting from January 1978 to December 2013. The bar plots suggest that both counts follow Zero inflated Poisson distribution, whereas the ACFs indicate that the counts are serially dependent. Finally, to illustrate the superiority of the proposed method they compare the method with zero-inflated integer-valuedautoregressive (ZIINAR) models.

3.2 Bivariate Copula-Based Model for Count Time Series Data

Copula based bivariate model

Suppose we have $\{Y_{1t}\}$ and $\{Y_{2t}\}$ jointly observed at timepoints t=1, 2, …, n, with the assumption that each series $\{Y_{1t}\}$ and $\{Y_{2t}\}$ follows a copula-based Markov process described on section 3.1. Let's mean vector, correlation matrix of the bivariate series as μ_t

and $\tau(t, t-1)$ which are described as below.

$$\mu_t = E(Y_t) = \begin{bmatrix} E(Y_{1t}) \\ E(Y_{2t}) \end{bmatrix}$$

$$\tau(t, t-1) = COV(Y_t, Y_{t-1}) \begin{bmatrix} COV(Y_{1t}, Y_{1,t-1}) & COV(Y_{1t}, Y_{2,t-1}) \\ COV(Y_{2t}, Y_{1,t-1}) & COV(Y_{2t}, Y_{2,t-1}) \end{bmatrix}$$

Here the diagonal elements of the matrix represent the serial dependence between two series, while the off-diagonal elements describe the cross-correlation between two time series.

The joint distribution of Y_{1t} and Y_{2t} given Y_{1t-1}, Y_{2t-1} for t=1, 2,…,n is given by

$$\mathrm{f}(y_{1t}, y_{2t} | y_{1t-1}, y_{2t-1}) = \int_{V^{-1}(F_{1,t}^{-})}^{V^{-1}(F_{1,t}^{+})} \int_{V^{-1}(F_{2,t}^{-})}^{V^{-1}(F_{2,t}^{+})} V_2(z_1, z_2, R) d z_2 \, dz_1$$

where V^{-1} is either the inverse cdf (Cumulative distribution function) of the normal distribution or the t-distribution with $V_2(., R)$ being the bivariate normal or t-distribution, respectively. R is correlation matrix capturing the cross correlation between two time series which is described below.

$$\mathrm{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

The limits of the bivariate integral can be calculated as below.

$F_{i,t}^{+} = F(y_{i,t} | y_{i,t-1})$ *and* $F_{i,t}^{-} = F(y_{i,t} - 1 | y_{i,t-1})$, for i= 1,2 where,

$$F(y_{i,t} | y_{i,t-1}) = \frac{F_{12}(y_{i,t}, y_{i,t-1}) - F_{12}(y_{i,t}, y_{i,t-1} - 1)}{f_{t-1}(y_{i,t-1}; \boldsymbol{\theta})}$$

and

$$F_{12}(y_{i,t}, y_{i,t-1}) = C(F_t(y_{i,t}), F_{t-1}(y_{i,t-1} - 1); \delta)$$

$C(.; \delta)$ represents the bivariate copula function with dependence parameter δ, describing the serial dependence in a single series, and $\boldsymbol{\theta}$ is a vector of the marginal parameters.

Likelihood and parameter estimation for the bivariate model

Likelihood based inference were conducted with maximizing the log-likelihood function of the bivariate distribution. The corresponding likelihood function for the joint distribution is given by,

$$L(\vartheta, y) = f(Y_{11}, Y_{21}). \prod_{t=2}^{n} f(Y_{1t}, Y_{2t} | Y_{1t-1}, Y_{2t-1}) \quad (4)$$

Where $\vartheta = (\theta', \delta_1, \delta_2, \rho)'$,where θ is the marginal parameter vector and δ_1, δ_2 are parameters associated with the serial dependence in each time series respectively. The cross correlation between the two-time series is captured by ρ.

We can construct the log -likelihood function $l(\vartheta, y)$ as below.

$$l(\vartheta, y) = \log(f(Y_{1t}, Y_{2t})) + \sum_{t=2}^{n} \log f(Y_{1t}, Y_{2t} | Y_{1t-1}, Y_{2t-1}).$$

The likelihood function ($l(\vartheta, y)$) contains either a bivariate normal or t-integral function which unable us to use the standard maximization procedures to get the ML estimates. Due to this reason, we evaluated the bivariate integral function using the standard randomized importance sampling method.

We present simulation results for the proposed bivariate model in Section 3.1 after expanding from univariate to bivariate model. For each univariate time series, we considered a copula-based Markov model, where a copula family was used for the joint distribution of subsequent observations, and then, coupled these two-time series using another copula at each time point.

The parameters of the marginal Poisson distribution are shown in Table 3 and Table 4 for positive and negative cross correlations, respectively. Here $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$ denote the means, $\boldsymbol{\omega_1}$ and $\boldsymbol{\omega_2}$ denote zero inflation parameters, $\boldsymbol{\delta_1}$ and $\boldsymbol{\delta_2}$ denote the serial dependence of marginal distributions. ρ is measure of the cross correlation between the two time series distributions.

The Gaussian copula was used to construct marginal distributions for 300 replicates with sample sizes of 100,300 ,500 and the true parameter values are presented in brackets. The count time series with positive cross correlation is presented in Figure 1, and the joint density is shown in Figure 2. When observing the parameter estimates displayed in Table 3, we can state that the estimated values are more precise and converges to the true parameter values as the sample size increases.

Bivariate ZI count time series models

Table 2. Parameter estimates for the bivariate ZI Poisson model with positive cross correlation

Sample Size	Parameters	Estimate	SE	MSE	MAE
100	λ_1 (3)	3.4021	0.3887	0.3123	0.4599
	ω_1(0.3)	0.3333	0.0835	0.0081	0.0701
	λ_2(5)	5.1993	0.3832	0.1860	0.3337
	ω_2(0.4)	0.4026	0.0686	0.0047	0.0537
	δ_1(0.6)	0.5425	0.0837	0.0103	0.0788
	δ_2(0.4)	0.3628	0.0963	0.0106	0.0806
	ρ(0.5)	0.4822	0.0911	0.0086	0.0748
300	λ_1 (3)	3.4051	0.1974	0.2030	0.4082
	ω_1(0.3)	0.3380	0.0447	0.0034	0.0471
	λ_2(5)	5.1816	0.2097	0.0768	0.2226
	ω_2(0.4)	0.4065	0.0386	0.0015	0.0309
	δ_1(0.6)	0.5540	0.0433	0.0040	0.0524
	δ_2(0.4)	0.3669	0.0544	0.0040	0.0492
	ρ(0.5)	0.4711	0.0493	0.0033	0.0441
500	λ_1 (3)	3.4105	0.1721	0.1980	0.4108
	ω_1(0.3)	0.3408	0.0365	0.0030	0.0456
	λ_2(5)	5.1843	0.1622	0.0602	0.2028
	ω_2(0.4)	0.4084	0.0293	0.0009	0.0246
	δ_1(0.6)	0.5558	0.0320	0.0030	0.0465
	δ_2(0.4)	0.3700	0.0430	0.0027	0.0413
	ρ(0.5)	0.4720	0.0392	0.0023	0.0379

Source: Fernando, D., Alqawba, M., Fernando, D., Diawara, N.& Samad, M. (2022).

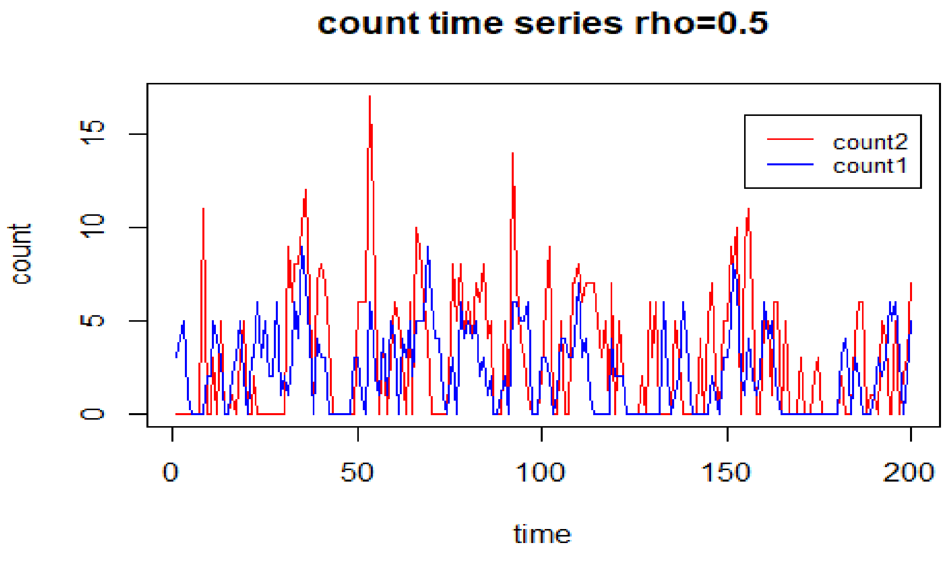

Figure 1. Plot of individual ZI count time series with positive cross-correlation

Source: Fernando, D., Alqawba, M., Fernando, D., Diawara, N.& Samad, M. (2022).

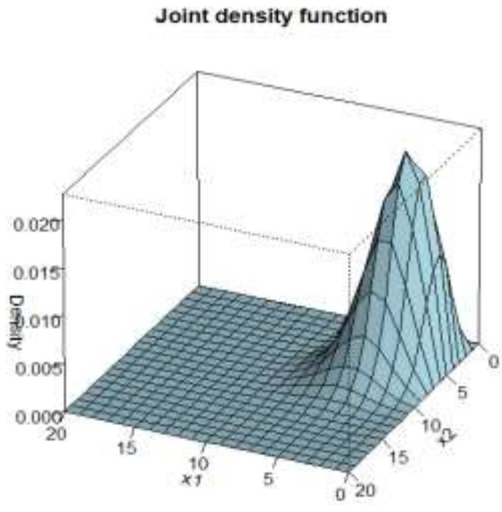

Figure 2. Joint probability density function for the bivariate ZI model with positive cross-correlation

Source: Fernando, D., Alqawba, M., Fernando, D., Diawara, N.& Samad, M. (2022).

There are times when the correlation is negative and table 4 shows the parameter estimates for such scenarios. The Gaussian copula was again used in constructing marginal distributions for 300 replicates with sample sizes of 100, 300, 500 and the true parameter values are presented in brackets. The count time series with negative cross correlation is illustrated in Figure 3, and the joint density is shown in Figure 4. The estimated parameters in Table 4 are more precise and converge to the true parameter values with increasing sample size as observed before.

These results are new because a large body of the literature focuses on positive correlations. Therefore, our proposed algorithm can handle less restrictive cases of ZI count time series data.

Table 3. Parameter estimates for the bivariate ZI Poisson model with negative cross correlation

Sample Size	Parameters	Estimate	St_Dev	MSE	MAE
100	λ_1 (3)	3.417	0.388	0.324	0.474
	ω_1(0.3)	0.341	0.084	0.074	0.074
	λ_2(5)	5.225	0.382	0.196	0.354
	ω_2(0.4)	0.408	0.070	0.056	0.056
	δ_1(0.6)	0.549	0.085	0.010	0.077
	δ_2(0.4)	0.368	0.103	0.012	0.086
	ρ(-0.4)	-0.391	0.104	0.011	0.081
300	λ_1 (3)	3.4072	0.2016	0.2063	0.4110
	ω_1(0.3)	0.3378	0.0455	0.0035	0.0477
	λ_2(5)	5.2100	0.1965	0.0826	0.2331
	ω_2(0.4)	0.4077	0.0379	0.0015	0.0313
	δ_1(0.6)	0.5529	0.0458	0.0043	0.0534
	δ_2(0.4)	0.3683	0.0537	0.0039	0.0499
	ρ(-0.4)	-0.3815	0.0559	0.0035	0.0465
500	λ_1 (3)	3.4181	0.1727	0.2045	0.4182
	ω_1(0.3)	0.3364	0.0348	0.0025	0.0412
	λ_2(5)	5.1984	0.1575	0.0641	0.2138
	ω_2(0.4)	0.4094	0.0304	0.0010	0.0254
	δ_1(0.6)	0.5524	0.0321	0.0033	0.0493
	δ_2(0.4)	0.3731	0.0417	0.0025	0.0388
	ρ(-0.4)	-0.3794	0.0460	0.0025	0.0414

Source: Fernando, D., Alqawba, M., Fernando, D., Diawara, N.& Samad, M. (2022).

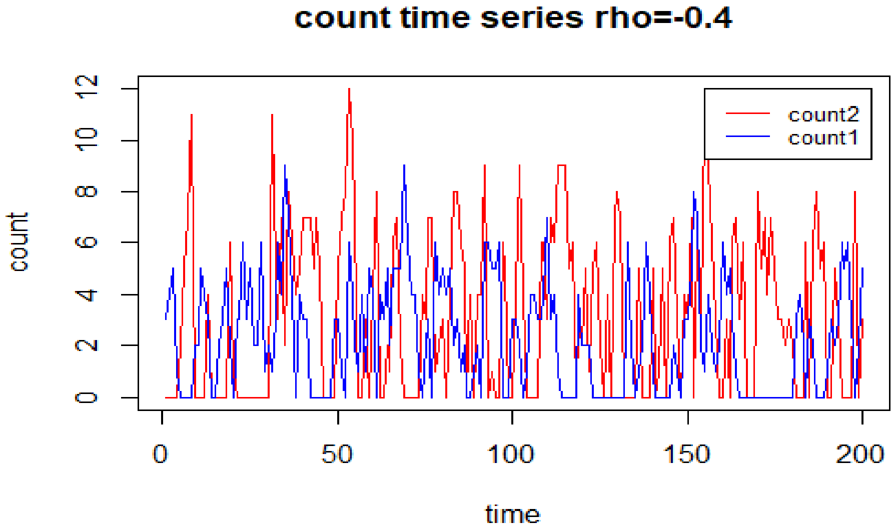

Figure 3. Plot of individual ZI count time series data with negative cross-correlation

Source: Fernando, D., Alqawba, M., Fernando, D., Diawara, N.& Samad, M. (2022)

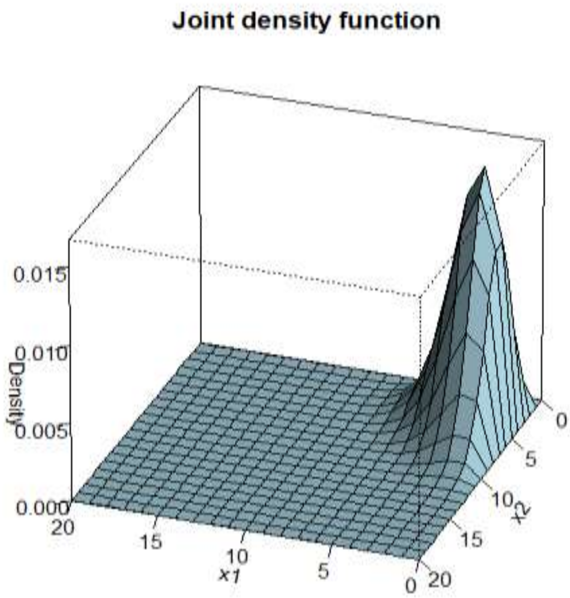

Figure 4. Joint probability density function for the bivariate ZI model with negative cross correlation

Source: Fernando, D., Alqawba, M., Fernando, D., Diawara, N.& Samad, M. (2022).

Applications

The proposed class of method can be applied to model bivariate zero inflated count time series data in the presence of both temporal dependence and cross correlation.

Wang et al. (2013) proposed a bivariate zero inflated poison model to analyze occupational injuries. Alqawba et al. (2021) applied this framework to model monthly counts of forgery and fraud in the 61st police car beat in Pittsburgh, PA. Two count time series were selected to fit the proposed bivariate Poisson class of models under the clear evidence of the presence of serial dependence and cross correlation.

4. Extensions of the Bivariate Copula for Count Time Series Data

Many copulas have been proposed in the literature for the bivariate and multivariate distributions. The choice of the copula is mainly dictated by the dependence structure.

As shown in Größer and Okhrin (2021), the research on time series dependence and copula direction is productive and has numerous applications. They showed examples of bivariate copulas. Count time series data are observed in several applied disciplines such as environmental science, biostatistics, economics, public health, and finance. Sometimes, a specific count, usually zero, may occur more often than other counts. Moreover, overlooking the frequent occurrence of zeros could result in misleading inferences. A copula-based time series regression model for zero-inflated counts is developed. Applying ordinary Poisson and Negative Binomial distributions to these time series of counts may not be appropriate due to the frequent occurrence of zeros. A new form of ZI is called the Conway-Maxwell Poisson (CMP).

Alqawba et al. (2021) have extended the work done by Masarotto (2012) to include a class of models that accounts for ZI. The marginals are assumed to follow one of the ZIP, ZINB, and ZICMP distributions, and the serial dependence was modeled by a Gaussian copula with a correlation matrix that of a stationary ARMA process. Likelihood inference was carried out using sequential importance sampling. Simulated studies were conducted to evaluate the parameter estimation procedures. Model description and parameter misspecification or unidentifiability are always concerns from the data generation to real data analysis (Faugeras, 2017). Model assessment to check the goodness of fit for the proposed models was done via residual analysis. The proposed models were applied to the occupational health data. According to the residual analysis, the model fits the data adequately, but both ZINB and ZICMP seem to have a slight advantage over ZIP distribution. Future direction is to consider different model construction methods from the marginal regression such as Markov models to handle zero-inflated count time series data. Recently, the use of copula-based time series for ZI counts in the presence of covariates has been proposed in Alqawba et al. (2019) and Alqawba and Diawara (2020). The work considered the cases of ZIP, ZINB, and ZICMP distributed marginals. Likelihood-based inference is considered under a sequential sampling method to estimate both the marginal regression parameters and copula parameters. Improvements in the Bayesian Information Criteria were noted, as discussed in Joe (2014) and Dalla Valle et al. (2018). The applications of these models include occupational injury counts, arson counts, and sandstorm counts.

5. Further Developments and Conclusion

Several high-dimensional copulas are obtained from the bivariate version seen in the previous section. The bivariate time series copula becomes then very important. The vine copula is built from blocks of bivariate version of higher dimension (Acar et al. 2019, Czado). We will only mention the Hierachical Archimedean copula, the Multivariate Archimax copula, the Factor copula, and the Vine copula. Copula functions are particularly interesting in capturing dependence with pairwise Kendall's correlations for invariance to monotonic transformations of marginal distributions. The copula is Archemedian and is applicable for higher than bivariate dimensions of the correlation between marginals (McNeil and Nešlehová, 2009). There is research on the symmetry of copula, and the family of measures under non-degenerate asymptotic distributions (Quessy and Bahraoui, 2018). The disentangling of features with copula transformation is also gaining popularity in so called deep Information bottleneck (DIB) to yield higher convergence rates (Wieczorek et al. 2018, Wieczorek and Roth 2020). As a measure, the copula can be thought as a transformation on a set, which is also a measure preserving transformation. Copulas are also obtained under non-monotonic transformations. Bardossy and Li (2008) proposed a v-transformed copula.

The ideas of Levy processes modelled via copula offer many areas of research (Liu et al., 2021).

The spatio-temporal dependence will become more of a priority as the research evolves. See more in Krupskii and Genton (2017). Bivariate time varying copulas are proposed in Acar et al. (2019). The dynamic vine copula is also adapted to the Bayesian inference (Kreuzer and Czado, 2019).

In this review, we have shown statistical and computational methods for bivariate count time series data analyses using copula distributions. The general framework for discrete count data and the bivariate nature of data are presented. The copula structure is described with details on its analytic perspectives. The identifiability and the choice of copula are very challenging in any discrete data setting and in the case of negative associations between components. As mentioned in Genest et al. (2011), Faugeras (2017) and in Trivedi and Zimmer (2007, 2017), the copula may not generate the perfect data distributions. Such concern is also pointed out in Durante and Sempi (2016). Copula can model bivariate dependence that are invariant under monotonic transformation only (Größer and Okhrin, 2021). When the dependence is weak, the FGM copula offers great alternative, but determining the most appropriate type of FGM copula to fit data is an open problem. Trivedi and Zimmer (2017) proposed several simulations to show these concerns.

Similar to any other functions, the copula functions cannot be deemed as the solution to all data problems. However, they offer a valuable alternative, especially in the case of discrete data. The research on discrete time series data is more important in this class of functions, especially for bivariate cases as the characterization of bivariate count dependence structure provides tools for may applied problems.

Conflict of Interest

We attest that the manuscript titled "Review of copula for bivariate distributions of zero-inflated count time series data" is original and has not been submitted to or considered for publication elsewhere. The authors declare that they have no competing or conflicts of interest with regard to this publication.

Funding Information

This study received no external funding.

References

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, *44*(2), 182-198. https://doi.org/10.1016/j.insmatheco.2007.02.001

Acar, E. F., Czado, C., & Lysy, M. (2019). Flexible dynamic vine copula models for multivariate time series data. *Econometrics and Statistics*, *12*, 181-197. https://doi.org/10.1016/j.ecosta.2019.03.002

Alqawba, M., & Diawara, N. (2021). Copula-based Markov zero-inflated count time series models with application. *Journal of Applied Statistics*, *48*(5), 786-803. https://doi.org/10.1080/02664763.2020.1748581

Alqawba, M., Diawara, N., & Chaganty, R. (2019). Zero-Inflated Count Time Series Models Using Gaussian Copula. *Sequential Analysis Journal: Design methods and Applications*, *38*(3), 342–357. https://doi.org/10.1080/07474946.2019.1648922

Alqawba, M., Fernando, D., & Diawara, N. (2021). A Class of Copula-Based Bivariate Poisson Time Series Models with Applications. *Computation*, *9*(10), 108. https://doi.org/10.3390/computation9100108

Armillotta, M., & Fokianos, K. (2021). Poisson network autoregression. *arXiv preprint arXiv:2104.06296*.

Bahraoui, T., Bouezmarni, T., & Quessy, J. F. (2018). Testing the symmetry of a dependence structure with a characteristic function. *Dependence Modeling*, *6*(1), 331-355. https://doi.org/10.1515/demo-2018-0019

Bedford, T., & Cooke, R. M. (2002). Vines--a new graphical model for dependent random variables. *The Annals of Statistics*, *30*(4), 1031-1068. https://doi.org/10.1214/aos/1031689016

Cook, R. D., & Johnson, M. E. (1981). A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *43*(2), 210-218. https://doi.org/10.1111/j.2517-6161.1981.tb01173.x

Czado, C. (2019). Analyzing dependent data with vine copulas. *Lecture Notes in Statistics*, *Springer*, *222*. https://doi.org/10.1007/978-3-030-13785-4

Czado, C., Schepsmeier, U., & Min, A. (2012). Maximum likelihood estimation of mixed C-vines with application to exchange rates. *Statistical Modelling*, *12*(3), 229-255. https://doi.org/10.1177/1471082X1101200302

Davis, R. A., Fokianos, K., Holan, S. H., Joe, H., Livsey, J., Lund, R., ... & Ravishanker, N. (2021). Count time series: A methodological review. *Journal of the American Statistical Association*, *116*(535), 1533-1547. https://doi.org/10.1080/01621459.2021.1904957

Davis, R. A., Holan, S. H., Lund, R., & Ravishanker, N. (Eds.). (2016). *Handbook of discrete-valued time series*. CRC Press. https://doi.org/10.1201/b19485

Deng, Y., & Chaganty, N. R. (2021). Pair-copula models for analyzing family data. *Journal of Statistical Theory and Practice*, *15*(1), 1-12. https://doi.org/10.1007/s42519-020-00146-z

Durante, F., & Sempi, C. (2016). *Principles of copula theory* (Vol. 474). Boca Raton, FL: CRC press. https://doi.org/10.1201/b18674

Durante, F., Sánchez, J. F., & Sempi, C. (2018). A note on bivariate Archimax copulas. *Dependence Modeling*, *6*(1), 178-182. https://doi.org/10.1515/demo-2018-0011

Fatahi, A. A., Noorossana, R., Dokouhaki, P., & Moghaddam, B. F. (2012). Copula-based bivariate ZIP control chart for monitoring rare events. *Communications in Statistics-Theory and Methods*, *41*(15), 2699-2716. https://doi.org/10.1080/03610926.2011.556296

Fokianos, K. (2021). Multivariate count time series modelling. *Econometrics and Statistics*. https://doi.org/10.1016/j.ecosta.2021.11.006

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, *74*(3), 549-555. https://doi.org/10.1093/biomet/74.3.549

Genest, C., & MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, *40*(4), 280-283. https://doi.org/10.1080/00031305.1986.10475414

Genest, C., Nešlehová, J., & Ziegel, J. (2011). Inference in multivariate Archimedean copula models. *Test*, *20*(2), 223-256. https://doi.org/10.1007/s11749-011-0250-6

Gräler, B. (2014). Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, *10*, 87-102. https://doi.org/10.1016/j.spasta.2014.01.001

Größer, J., & Okhrin, O. (2022). Copulae: An overview and recent developments. *Wiley Interdisciplinary Reviews: Computational Statistics*, *14*(3), e1557. https://doi.org/10.1002/wics.1557

Gumbel, E. J. (1958). Distributions à plusieurs variables dont les marges sont données. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences*, *246*(19), 2717-2719.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, *56*(4), 1030-1039. https://doi.org/10.1111/j.0006-341X.2000.01030.x

Han, Z., & De Oliveira, V. (2016). On the correlation structure of Gaussian copula models for geostatistical count data. *Australian & New Zealand Journal of Statistics*, *58*(1), 47-69. https://doi.org/10.1111/anzs.12140

Han, Z., & De Oliveira, V. (2020). Maximum likelihood estimation of Gaussian copula models for geostatistical count data. *Communications in Statistics-Simulation and Computation*, *49*(8), 1957-1981. https://doi.org/10.1080/03610918.2018.1508705

Irannezhad, E., Prato, C. G., Hickman, M., & Mohaymany, A. S. (2017). Copula-based joint discrete–continuous model of road vehicle type and shipment size. *Transportation Research Record*, *2610*(1), 87-96. https://doi.org/10.3141/2610-10

Joe, H., Li, H., & Nikoloulopoulos, A. K. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, *101*(1), 252-270. https://doi.org/10.1016/j.jmva.2009.08.002

Johnson, M. E., & Tenenbein, A. (1981). A bivariate distribution family with specified marginals. *Journal of the American Statistical Association*, *76*(373), 198-201. https://doi.org/10.1080/01621459.1981.10477628

Karlis, D., & Pedeli, X. (2013). Flexible bivariate INAR (1) processes using copulas. *Communications in Statistics-Theory and Methods*, *42*(4), 723-740. https://doi.org/10.1080/03610926.2012.754466

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1-14. https://doi.org/10.2307/1269547

Lin, H., & Chaganty, N. R. (2021). Multivariate distributions of correlated binary variables generated by pair-copulas. *Journal of Statistical Distributions and Applications*, *8*(1), 1-14. https://doi.org/10.1186/s40488-021-00118-z

Liu, Y., Djurić, P. M., Kim, Y. S., Rachev, S. T., & Glimm, J. (2021). Systemic risk modeling with lévy copulas. *Journal of Risk and Financial Management*, *14*(6), 251. https://doi.org/10.3390/jrfm14060251

Ma, Z., Hanson, T. E., & Ho, Y. Y. (2020). Flexible bivariate correlated count data regression. *Statistics in Medicine*, *39*(25), 3476-3490. https://doi.org/10.1002/sim.8676

Masarotto, G. (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics*, *6*, 1517-1549. https://doi.org/10.1214/12-EJS721

McNeil, A. J., & Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and ℓ1-norm symmetric distributions. *The Annals of Statistics*, *37*(5B), 3059-3097. https://doi.org/10.1214/07-AOS556

Nikoloulopoulos, A. K., & Karlis, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics*, *37*(9), 1555-1568. https://doi.org/10.1080/02664760903093591

Nikoloulopoulos, A. K., & Moffatt, P. G. (2019). Coupling couples with copulas: analysis of assortative matching on risk attitude. *Economic Inquiry*, *57*(1), 654-666. https://doi.org/10.1111/ecin.12726

Panagiotelis, A., Czado, C., & Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, *107*(499), 1063-1072. https://doi.org/10.1080/01621459.2012.682850

Rao, M. B., & Subramanyam, K. (1990). The structure of some classes of bivariate distributions and some applications. *Computational Statistics & Data Analysis*, *10*(2), 175-187. https://doi.org/10.1016/0167-9473(90)90063-N

Ridout, M., Hinde, J., & Demétrio, C. G. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, *57*(1), 219-223. https://doi.org/10.1111/j.0006-341X.2001.00219.x

Rigby, R. A., Stasinopoulos, M. D., Heller, G. Z., & De Bastiani, F. (2019). *Distributions for modeling location, scale, and shape: Using GAMLSS in R*. CRC press. https://doi.org/10.1201/9780429298547

Safari-Katesari, H., Samadi, S. Y., & Zaroudi, S. (2020). Modelling count data via copulas. *Statistics*, *54*(6), 1329-1355. https://doi.org/10.1080/02331888.2020.1867140

Sellers, K. F., & Raim, A. (2016). A flexible zero-inflated model to address data dispersion. *Computational Statistics & Data Analysis*, *99*, 68-80. https://doi.org/10.1016/j.csda.2016.01.007

Shamma, N., Mohammadpour, M., & Shirozhan, M. (2020). A time series model based on dependent zero inflated counting series. *Computational Statistics*, *35*(4), 1737-1757. https://doi.org/10.1007/s00180-020-00982-4

Shi, P., & Zhang, W. (2015), Private information in healthcare utilization: specification of a copula-based hurdle model. *J. R. Stat. Soc. A*, *178*, 337-361. https://doi.org/10.1111/rssa.12065

Sklar, A. (1973). Random variables, joint distribution functions and copulas. *Kybernetika*, *9*, 449–460.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, *8*, 229-231.

Trivedi, P. K., & Zimmer, D. M. (2007). *Copula Modeling: An Introduction for Practitioners*. Foundations and Trends in Econometrics. https://doi.org/10.1561/0800000005

Trivedi, P., & Zimmer, D. (2017). A note on identification of bivariate copulas for discrete count data. *Econometrics*, *5*(1), 10. https://doi.org/10.3390/econometrics5010010

Valle, L. D., Leisen, F., & Rossini, L. (2018). Bayesian non-parametric conditional copula estimation of twin data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *67*(3), 523-548. https://doi.org/10.1111/rssc.12237

van den Heuvel, E. R., van Driel, S. A., & Zhan, Z. (2022). A bivariate zero-inflated Poisson control chart: Comments and corrections on earlier results. *Communications in Statistics-Theory and Methods*, *51*(10), 3438-3445. https://doi.org/10.1080/03610926.2020.1736304

Wang, K., Lee, A. H., Yau, K. K., & Carrivick, P. J. (2003). A bivariate zero-inflated Poisson regression model to analyze occupational injuries. *Accident Analysis & Prevention*, *35*(4), 625-629. https://doi.org/10.1016/S0001-4575(02)00036-2

Weiß, C. H., Möller, T., & Kim, H. Y. (2020). Modelling counts with state-dependent zero inflation. *Statistical modelling*.

Wieczorek, A., & Roth, V. (2020). On the difference between the information bottleneck and the deep information bottleneck. *Entropy*, *22*(2), 131. https://doi.org/10.3390/e22020131

Wieczorek, A., Wieser, M., Murezzan, D., & Roth, V. (2018). Learning sparse latent representations with the deep copula information bottleneck. *arXiv preprint arXiv:1804.06216*.

Yang, Q., Xu, M., Lei, X., Zhou, X., & Lu, X. (2014). A Methodological Study on AMH Copula-Based Joint Exceedance Probabilities and Applications for Assessing Tropical Cyclone Impacts and Disaster Risks (Part I). *Tropical Cyclone Research and Review*, *3*(1), 53-62.

Young, D. S., Roemmele, E. S., & Shi, X. (2022). Zero-inflated modeling part II: Zero-inflated models for complex data structures. *Wiley Interdisciplinary Reviews: Computational Statistics*, *14*(2), e1540.

Yu, R., Yang, R., Zhang, C., Špoljar, M., Kuczyńska-Kippen, N., & Sang, G. (2020). A vine copula-based modeling for identification of multivariate water pollution risk in an interconnected river system network. *Water*, *12*(10), 2741. https://doi.org/10.3390/w12102741

Zhang, Y., & Lam, J. S. L. (2016). A copula approach in the point estimate method for reliability engineering. *Quality and Reliability Engineering International*, *32*(4), 1501-1508. https://doi.org/10.1002/qre.1860