Copula-Based Models For ZI Count Time Series Data

Dimuthu Fernando and Norou Diawara
Department of Mathematics and Statistics
Old Dominion University

Introduction

- Count time series data are found in multiple applications. Such time series counts come with inflation and in a bivariate form that captures not only serial dependence within each time series but also interdependence between the two series.
- We constructed a class of bivariate integer-valued time series model [1] by applying either the bivariate Gaussian copula or the bivariate t copula functions.
- This method allows for flexible dependence structures between the two time series and within each one.
- First order Markov chains are considered with zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB) and zero-inflated Conway-Maxwell-Poisson (ZICMP) marginals.

Simulation Results

Gaussian copula and Student's t copula were selected as candidate copula families with true parameters (λ_1=3, λ_2=5, ω_1=0.3, ω_2=0.4, δ_1=0.6, δ_2=0.4, ρ=0.5).

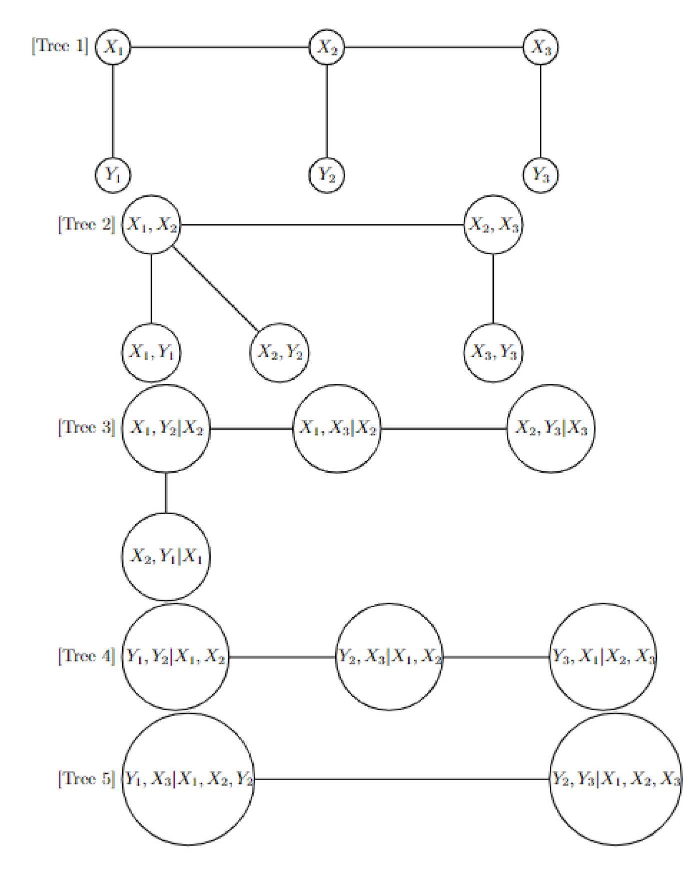
Sample Size	Parameters	Estimate	SE	MSE	MAE
	$\lambda_1(3)$	3.402	0.389	0.312	0.460
	$\omega_1(0.3)$	0.333	0.083	0.008	0.070
	$\lambda_2(5)$	5.199	0.383	0.186	0.334
100	$\omega_2(0.4)$	0.403	0.069	0.005	0.054
	$\delta_1(0.6)$	0.542	0.084	0.010	0.079
	$\delta_2(0.4)$	0.363	0.096	0.011	0.081
	$\rho(0.5)$	0.482	0.091	0.009	0.075
	$\lambda_1(3)$	3.405	0.197	0.203	0.408
	$\omega_1(0.3)$	0.338	0.045	0.003	0.047
	$\lambda_2(5)$	5.182	0.210	0.077	0.223
300	$\omega_2(0.4)$	0.406	0.039	0.002	0.031
	$\delta_1(0.6)$	0.554	0.043	0.004	0.052
	$\delta_2(0.4)$	0.367	0.054	0.004	0.049
	$\rho(0.5)$	0.471	0.049	0.003	0.044
	$\lambda_1(3)$	3.410	0.172	0.198	0.411
	$\omega_1(0.3)$	0.341	0.037	0.003	0.046
	$\lambda_2(5)$	5.184	0.162	0.060	0.203
500	$\omega_2(0.4)$	0.408	0.029	0.001	0.025
	$\delta_1(0.6)$	0.556	0.032	0.003	0.047
	$\delta_2(0.4)$	0.370	0.043	0.003	0.041
	$\rho(0.5)$	0.472	0.039	0.002	0.038

Sample Size	Parameters	Estimate	St_Dev	MSE	MAE
	$\lambda_1(3)$	2.634	0.396	0.464	0.550
	$\omega_1(0.3)$	0.338	0.061	0.008	0.071
	$\kappa_1(0.4)$	0.339	0.051	0.008	0.071
	$\lambda_2(5)$	4.768	1.004	3.021	1.289
100	$\omega_2(0.4)$	0.412	0.055	0.007	0.065
	$\kappa_2(0.6)$	0.577	0.072	0.015	0.010
	$\delta_1(0.6)$	0.576	0.071	0.016	0.101
	$\delta_2(0.4)$	0.379	0.091	0.026	0.128
	$\rho(0.5)$	0.488	0.086	0.019	0.104
	$\lambda_1(3)$	2.813	0.234	0.234	0.437
	$\omega_1(0.3)$	0.319	0.038	0.003	0.046
	$\kappa_1(0.4)$	0.364	0.031	0.005	0.006
	$\lambda_2(5)$	4.825	0.687	0.501	0.559
300	$\omega_2(0.4)$	0.413	0.032	0.001	0.028
	$\kappa_2(0.6)$	0.577	0.050	0.003	0.044
	$\delta_1(0.6)$	0.590	0.042	0.002	0.034
	$\delta_2(0.4)$	0.387	0.052	0.003	0.042
	$\rho(0.5)$	0.493	0.052	0.003	0.042
	$\lambda_1(3)$	2.816	0.209	0.224	0.432
	$\omega_1(0.3)$	0.318	0.031	0.002	0.039
	$\kappa_1(0.4)$	0.368	0.028	0.005	0.067
	$\lambda_2(5)$	4.814	0.056	0.363	0.492
500	$\omega_2(0.4)$	0.411	0.028	0.001	0.024
	$\kappa_2(0.6)$	0.574	0.042	0.002	0.040
	$\delta_1(0.6)$	0.587	0.036	0.001	0.031
	$\delta_2(0.4)$	0.387	0.041	0.002	0.034
	$\rho(0.5)$	0.495	0.037	0.001	0.029

The COPAR model

- The idea is to extend the work of (Brechmann Czado, 2015) [2] to model zero-inflated count time series data.
- They introduced a copula-based model for stationary time series with a Markovian structure which captures both serial dependence and cross sectional dependence in multivariate time series.

Tree structure of 2 dimensional ($\{X_t\}$ and $\{Y_t\}$) COPAR model with 3 time points.



R-vine structure matrix (Morales, 2010) corresponding for the 2 dimensional COPAR model with 3 time points.

$$\begin{pmatrix} Y_3 & & & & & \\ Y_1 & X_3 & & & & \\ Y_2 & Y_1 & Y_2 & & & \\ X_1 & Y_2 & Y_1 & X_2 & & \\ X_2 & X_1 & X_1 & Y_1 & Y_1 & \\ X_3 & X_2 & X_2 & X_1 & X_1 & X_1 \end{pmatrix}$$

Matrix of copulas for the 2 dimensional COPAR model with 3 time points.

$$\begin{pmatrix} * & & & & & \\ C_2^Y & * & & & & \\ C_1^Y & C_2^{YX} & * & & & \\ C_2^{XY} & C_1^{YX} & C_1^Y & * & & \\ C_1^{XY} & C_2^X & C_1^{XY} & C_1^{YX} & * & \\ C_0^{XY} & C_1^X & C_0^{XY} & C_1^X & C_0^{XY} & * \end{pmatrix}$$

COPAR model of order k (COPAR(k)) is defined, where all pair copulas corresponding to a lag length greater than k are independence copulas.

$$C_{t-s}^X = C_{t-s}^{XY} = C_{t-s}^{YX} = C_{t-s}^Y = \Pi \quad for \quad t-s > k,$$

where Π denotes the independence copula, respectively.

Proposed Bivariate Model

- Consider the following series of $2-dimensional$ vector, $\{\boldsymbol{Y}_t\}_{t=1}^n$, where $\boldsymbol{Y}_t = (Y_{1t}, Y_{2t})'$ for $t = 1, 2, ..., n$.
- We assume that each series $\{Y_{1t}\}_{t=1}^n$ and $\{Y_{2t}\}_{t=1}^n$ follows a copula-based first order Markov process.
- The joint probability distribution of Y_{1t} and Y_{2t} given $Y_{1,t-1}$ and $Y_{2,t-1}$, respectively, for $t = 1, \ldots, n$, is given by :

$$f(y_{1t}, y_{2t} \,|y_{1,t-1}, y_{2,t-1}) = \int_{V^{-1}(F_{1,t}^-)}^{V^{-1}(F_{1,t}^+)} \int_{V^{-1}(F_{2,t}^-)}^{V^{-1}(F_{2,t}^+)} V_2(z_1, z_2, R) dz_2 dz_1$$

- V^{-1} denotes either the inverse cdf of the normal distribution or the t-distribution with $V_2(.,R)$ being the bivariate normal or t-distribution, respectively.
- R denotes the correlation matrix associated with the bivariate distribution capturing the cross-sectional dependence between the two series.
- The limits of the integral function are calculated using the the transition cdf of Y_{it} given $Y_{i,t-1}$, for $i = 1, 2$.
- The likelihood function for the proposed model is given by:

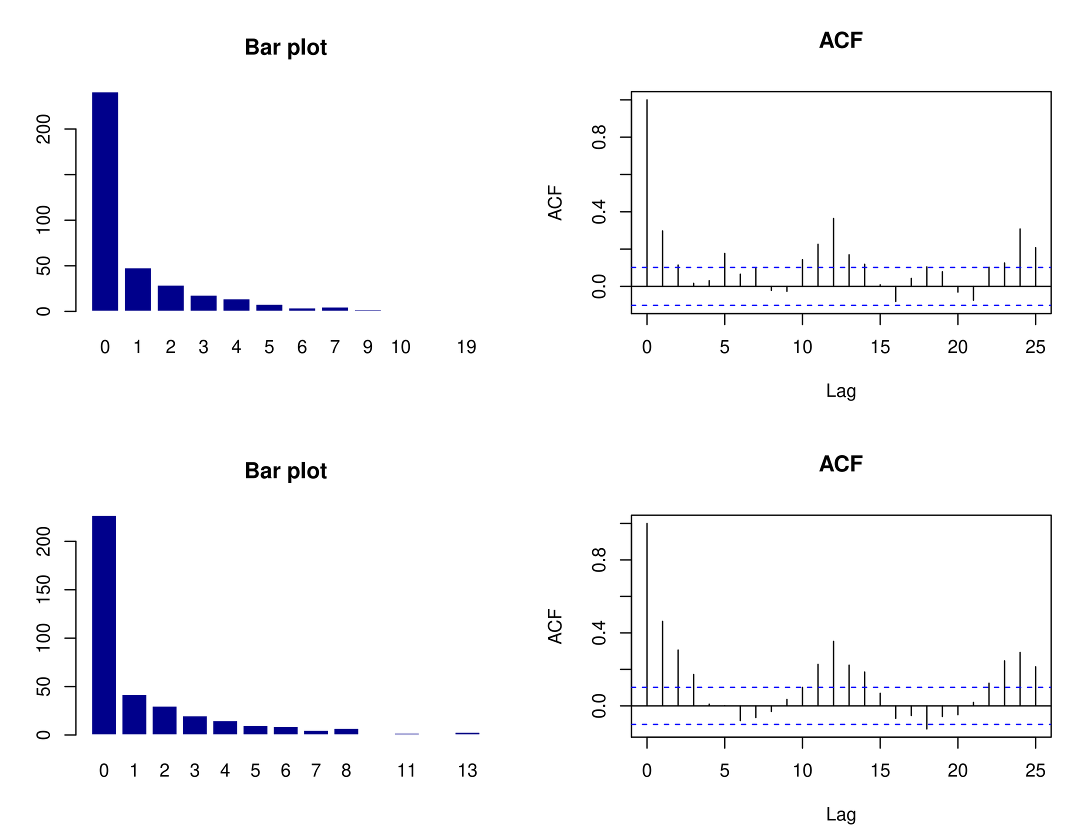
$$L(\boldsymbol{\vartheta}, \mathbf{y}) = f(y_{11}, y_{21}). \prod_{t=2}^n f(y_{1t}, y_{2t} \mid y_{1,t-1}, y_{2,t-1}),$$

where $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \delta_1, \delta_2, \rho)'$,is the marginal parameter vector, δ_1 and δ_2 are the copula parameters to deal with the first and second time series, respectively, and finally ρ the bivariate dependence parameter.

- When calculating ML estimates, the bivariate integral function was evaluated using the standard randomized importance sampling method.
- Taking the inverse of the Fisher information matrix, yields standard errors of the ML estimates.

Application

The data consists of the monthly counts of strong sandstorms recorded by the AQI airport stations in Eastern Province, Saudi Arabia.



Parameter estimates for the BZINB model.

Parameter	Estimate	SE
λ_1	1.758	0.197
ω_1	0.446	0.397
κ_1	0.961	0.419
λ_2	1.946	0.295
ω_2	0.348	0.727
κ_2	0.708	0.401
δ_1	0.412	0.059
δ_2	0.536	0.051
ρ	0.460	0.059

References

[1] Alqawba M., Fernando D., and Diawara N. A class of copula-based bivariate poisson time series models with applications. *Computation*, 108(9), 2021.

[2] Eike Christian Brechmann and Claudia Czado. Multivariate time series modeling using the copula autoregressive model. *Applied Stochastic Models in Business and Industry*, 31(4):495–514, 2015.