Statistica Sinica Preprint No: SS-2022-0406	
Title	Copula-based Analysis for Count Time Series
Manuscript ID	SS-2022-0406
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0406
Complete List of Authors	Mingze Zhang, Huixia Judy Wang and James Livsey
Corresponding Authors	Huixia Judy Wang
E-mails	judywang@gwu.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica

COPULA-BASED ANALYSIS FOR COUNT TIME SERIES

Mingze Zhang[1], Huixia Judy Wang[2] and James Livsey[3]

[1] *JPMorgan Chase & Co.*

[2] *Department of Statistics, The George Washington University*

[3] *U.S. Census Bureau*

Abstract:

We propose a new copula-based Markov model to analyze count time series data. One challenge in using copula to analyze count data is due to the identifiability issue that arises from the discrepancy of using a continuous copula function to characterize the discrete distribution. We find that identifiability can be ensured in the regression setup under one sufficient condition. Resolving the identifiability issue allows us to develop a method to select the appropriate copula to capture different types of temporal dependence, leading to more flexibility in modeling. We propose an estimation procedure and establish the asymptotic properties of the proposed estimators. For capturing temporal dependence, the proposed method is data-adaptive and computationally efficient. It also provides a convenient way to construct both point and interval predictions at a future time. Through a simulation study and the analysis of COVID-19 daily death data, we show that our method produces more stable point and interval predictions than existing methods based on Gaussian copula and autoregression.

Key words and phrases: Conditional quantile, copula, identifiability, Markov, temporal dependence

1. Introduction

With a surge in applications from finance, environmental science, and social science, there has been increasing interest in developing models and methods for analyzing time series of counts; see Davis and Liu (2016), Jung et al. (2006), Brandt et al. (2000), and Song et al. (2009). The analysis of count time series is challenging due to the nonuniqueness of the dependence measure caused by the existence of ties. There exist some methods in the literature but many of them either rely on restrictive distributional assumptions, are computationally intensive, or have limitations for capturing complex dependence. In this paper, we aim to develop a flexible and computationally convenient procedure with weaker assumptions that can accommodate various types of dependence.

Some existing methods assume that the data follow a Poisson or Negative Binomial distribution conditional on past observations or an intensity process; see e.g., Davis and Wu (2009), Davis et al. (2003), Davis et al. (2000), Fokianos and Tjøstheim (2011) and Fokianos et al. (2009). Ferland et al. (2006) proposed an integer-valued GARCH (INGARCH) model, and the observations $\{Y_t\}$ given the intensity process $\{\lambda_t\}$ follow a Poisson distribution and λ_t is a linear combination of its lagged values and lagged Y_t.

Other researchers have considered copula-based methods for analyzing time series of counts. One major advantage of copula-based methods is that it can separate the modeling of marginal and temporal dependence. Denuit and Lambert (2005) studied the concordance of integer-valued bivariate variables through the continuous extension (CE), and showed that the CE preserves Kendall's τ, a measure of rank correlation (Kendall, 1938). The CE works through constructing continuous variables by adding a random perturbation taking values in [0,1] to the discrete variables. The constructed continuous variables are also referred to as jittered data in the literature. Madsen (2009) and Madsen and Fang (2011) applied the CE idea to analyze discrete longitudinal and spatial data by using a Gaussian copula to capture the joint distribution of jittered data. Heinen and Rengifo (2007) proposed multivariate autoregressive models for count time series using CE and Gaussian copula. The

CE-based methods use jitters to fill in the gap between discrete values, and then estimate parameters by maximizing the surrogate likelihood of the jittered data. Nikoloulopoulos (2013, 2016) showed that the maximum surrogate likelihood estimators based on CE may lead to biased estimates for multivariate discrete data, since they use the information of jitters that are not in the observed data and the jittering is univariate without accounting for dependence. In addition, the CE-based methods are often time-consuming since the surrogate likelihood needs to be approximated by averaging over multiple jitters. More recently, Jia et al. (2021) developed a general count time series model class using Gaussian copula with Hermite expansions and particle filtering to approximate the likelihood.

The aforementioned copula-based methods all assume a latent Gaussian process, which may be restrictive in some applications. To gain more flexibility, we choose a copula from a class of copula families that better fits the data. However, this task faces a major challenge caused by the identifiability issue: the copula may not be unique when the marginal distribution is discrete. This issue arises from the discrepancy of using a continuous copula function to characterize the discrete distribution. To overcome the limitations of existing methods, we propose a new copula-based Markov model for analyzing count time series data. In this model framework, we utilize the copula to capture serial dependence, leveraging the Markov property, which is in the same spirit as considered in Chen and Fan (2006), Chen et al. (2009), Rémillard et al. (2012), and Tang et al. (2019) for continuous time series data. We formally address the identifiability issue and show that the identifiability can be ensured in the regression setup under one sufficient condition, that is, the dependence of the count response on one or more continuous covariates. Resolving the identifiability issue allows us to choose copula from a class of copula families to accommodate different types of temporal dependence, including, for instance, lower or upper tail dependence. We develop a maximum likelihood estimation procedure and establish the asymptotic properties of the proposed estimators. Compared to existing approaches for analyzing count time series, the proposed method is more flexible as it does not restrict to the latent Gaussian process, and it allows separate treatments of the univariate marginals and the temporal dependence. In addition, under the Markov assumption, we only need to model the dependence of data from adjacent time points. Therefore, the proposed procedure is computationally convenient as it avoids the challenge of high-dimensional integration as in Nikoloulopoulos (2013), Hughes (2015), Nikoloulopoulos (2016), to name a few.

The rest of this paper is organized as follows. In Section 2, we introduce the copula-based Markov model and study the identifiability issue. In addition, we present the maximum likelihood estimator, establish its asymptotic properties, and propose a copula selection procedure. We assess the numerical performance of the proposed method through a simulation study in Section 3 and the analysis of Covid-19 mortality data in Section 4. All technical proofs are provided in a separate supplementary file.

2. Proposed Method

2.1 Copula-based Markov model

Let $\boldsymbol{Y} = \{Y_t, t = 1, 2, \ldots, n\}$ be the observed count time serie data at n time points, and $\underline{\boldsymbol{X}}_n = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)^T$ be the corresponding $n \times d$ covariate matrix, where $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{td})^T$ for $t = 1, \ldots, n$ and d is the dimension of the covariates. For instance, in the coastal hurricane study, researchers are interested in predicting the annual hurricane counts by using covariates information, such as North Atlantic oscillation (NAO), Northwest

African rainfall, etc., from a few months earlier; see Elsner and Jagger (2006); Livsey et al. (2018).

We assume that the sequence of count $\{Y_t|\underline{\boldsymbol{X}}_t, t = 1, 2, \ldots, n\}$ is a Markov process of order p with discrete state space, that is, we assume

$$Y_t|Y_{t-1}, \ldots, Y_1, \underline{\boldsymbol{X}}_t \sim Y_t|Y_{t-1}, \ldots, Y_{t-p}, \underline{\boldsymbol{X}}_t. \tag{2.1}$$

Under this assumption, the probabilistic properties of the sequence are fully determined by the joint distribution of neighboring time points $Y_t, \ldots, Y_{t-p}$, denoted as $\mathcal{H}(y_t, \ldots, y_{t-p}|\underline{\boldsymbol{x}}_t)$. The Markov property (2.1) assumes that the current outcomes only depend on their recent past and corresponding covariates. We propose to model $\mathcal{H}(y_t, \ldots, y_{t-p}|\underline{\boldsymbol{x}}_t)$ through copula, the distribution function of a multivariate uniform distribution. Copulas are popular tools for multivariate analysis as they allow us to isolate the modeling of marginal distributions and temporal dependence structure; see a comprehensive review of copula in Nelsen (2006).

In the following, we first formally state the copula-based Markov model assumption for count time series. The Assumption A1 consists of three parts.

Assumption A1:

A1(a). At each time t, the marginal distribution of Y_t is characterized by $Y_t|\underline{\boldsymbol{X}}_t = \underline{\boldsymbol{x}}_t \sim G(y_t|\underline{\boldsymbol{x}}_t, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is the unknown marginal parameter vector.

A1(b). Given $\underline{\boldsymbol{X}}_t$, the process $\{Y_t, t = 1, \ldots, n\}$ is a stationary p-order Markovian satisfying $P(Y_t \le y|Y_{t-1}, \ldots, Y_1, \underline{\boldsymbol{X}}_t) = P(Y_t \le y|Y_{t-1}, \ldots, Y_{t-p}, \underline{\boldsymbol{X}}_t)$ for any $y \in \mathbb{R}^+$.

A1(c). There exists a $(p+1)$-dimensional copula C such that for any $y_t, \ldots, y_{t-p} \in \mathbb{R}_+$,

$$\mathcal{H}(y_t, \ldots, y_{t-p}|\underline{\boldsymbol{x}}_t) = C\{G(y_t|\underline{\boldsymbol{x}}_t; \boldsymbol{\beta}), \ldots, G(y_{t-p}|\underline{\boldsymbol{x}}_t; \boldsymbol{\beta}); \boldsymbol{\alpha}\},$$

where $C(\cdot;\boldsymbol{\alpha})$ is the copula function that does not vary with covariates, and $\boldsymbol{\alpha}$ is the unknown copula parameter vector. Furthermore, $C(\cdot;\boldsymbol{\alpha})$ is absolutely continuous with respect to Lebesque measure on $[0,1]^{p+1}$, and is neither the Fréchet-Hoeffding upper or lower bound.

Throughout we use $\boldsymbol{\beta}_0$ and $\boldsymbol{\alpha}_0$ to denote the true parameter values. Denote $U_t = G(Y_t|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta}_0)$. Suppose that $C(\cdot;\boldsymbol{\alpha})$ is the true copula function. Assumptions A1(b)-(c) imply that the process U_t is geometric ergodic and β-mixing for commonly used copulas such as Gaussian, Clayton, Gumbel, and t copulas. This can be shown by checking the sufficient conditions given in Proposition 2.1 of Chen and Fan (2006) and following similar arguments as in the proof of Theorem 2.1 in Chen et al. (2009).

For continuous time series, by Sklar's theorem (Sklar, 1959; Nelsen, 2006), we can represent $\mathcal{H}(y_t,\ldots,y_{t-p}|\underline{\boldsymbol{x}}_t)$ by the marginal conditional distribution function $G(\cdot|\underline{\boldsymbol{x}}_t,\boldsymbol{\beta})$ of Y_t and a unique copula function $C(\cdot)$. However, when the marginal distribution function $G(\cdot)$ is discrete for count data, the joint distribution $\mathcal{H}(\cdot)$ is uniquely defined only on the support of the margins. The discrepancy between the continuity of copula functions and discreteness of count data leads to the identifiability issue, that is, there may exist more than one copula satisfying condition A1(c). For now, we assume that $C(\cdot)$ is a prespecified copula function that satisfies A1(c). We will discuss how to overcome this identifiability issue in Section 2.2. For ease of presentation, we focus on $p=1$ hereafter, but the proposed method can be adapted for general p with some modification and additional computation.

2.2 Identifiability issue

Sklar (1959) suggested that the copula function is unique only over the domain of the marginal distributions. For continuous data, the marginal CDF takes value over the entire range of $[0, 1]$. However, for count data, the marginal CDF can only take countable values on [0,1], so the copula is not unique outside the support of the marginal CDF.

Several papers have discussed the identifiability issue of copula for discrete data. Genest and Nešlehová (2007) pointed out that the identifiability issue could be more damaging when the marginal distribution concentrates on a small number of values. This issue will tend to diminish when the discrete variable puts a positive probability on more and more possible outcomes. Trivedi and Zimmer (2017) conducted some simulations and concluded that copulas for discrete count outcomes fail to capture the dependence at extremely small means. Yang et al. (2020) discussed the identifiability issue and proposed a nonparametric estimation of copula functions under the regression setup when the outcome is discrete.

This paper focuses on copula-based analysis for count data in a regression setup. Under condition A1, the marginal distributions vary with covariates, so the inclusion of continuous covariates in the marginal can expand the region of support for copula identifiability. For example, in the Poisson regression model, $G(0|\underline{\boldsymbol{x}}_t, \boldsymbol{\beta}) = \exp\{-\exp(\underline{\boldsymbol{x}}_t{}'\boldsymbol{\beta})\}$. Hence, the inclusion of continuous covariates widens the range of $G(\cdot)$ from a discrete number of points to an interval spanning across $\underline{\boldsymbol{x}}_t$. This, combined with A1(c), ensures that the copula function can be uniquely determined over the region comprising possible values of $[0, 1]^p$.

Proposition 1. *Suppose the conditions in Assumption A1 are met. Additionally, assume that for any* $u \in (0, 1)$*, there exists a* $\mathbf{x}$ *and a value* z *that depends on* $\mathbf{x}$ *such that* $G(z|\mathbf{x}, \boldsymbol{\beta}) = u$*. Under these conditions, the copula function over* $[0, 1]^p$ *can be uniquely determined.*

The identifiablity condition in Proposition 1 is easily satisfied under the regression setup. For example, it is guaranteed when there is at least one continuous covariate with a sufficiently wide support and a nonzero coefficient in the marginal distribution $G(\cdot|\underline{\boldsymbol{x}}_t, \boldsymbol{\beta})$.

2.3 Parameter estimation

We present a maximum likelihood estimator for both the marginal parameter $\boldsymbol{\beta}$ and the copula parameter $\boldsymbol{\alpha}$. We first derive the likelihood function for the sequence $\{Y_t\}$ under the Markov property with order $p = 1$. From Assumption A1(c) and Song (2000), we can obtain the joint probability mass function (PMF) of (Y_{t-1}, Y_t) for any $t > 1$ as

$$\begin{aligned}
&P(Y_t = y_t, Y_{t-1} = y_{t-1}|\underline{\boldsymbol{X}}_t = \underline{\boldsymbol{x}}_t)\\
=&C(G(y_t|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta}), G(y_{t-1}|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta});\boldsymbol{\alpha}) - C(G(y_t|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta}), G(y_{t-1}-1|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta});\boldsymbol{\alpha})\\
&- C(G(y_t-1|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta}), G(y_{t-1}|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta});\boldsymbol{\alpha}) + C(G(y_t-1|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta}), G(y_{t-1}-1|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta});\boldsymbol{\alpha})\\
=&C(u_t, u_{t-1};\boldsymbol{\alpha}) - C(u_t, v_{t-1};\boldsymbol{\alpha}) - C(u_{t-1}, v_t;\boldsymbol{\alpha}) + C(v_t, v_{t-1};\boldsymbol{\alpha}),
\end{aligned} \tag{2.2}$$

where $u_t = G(y_t|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta})$ for $y_t = 0, 1, 2, \ldots$, and $v_t = G(y_t-1|\underline{\boldsymbol{x}}_t;\boldsymbol{\beta})$ for $y_t = 1, 2, \ldots$ and $v_t = 0$ for $y_t = 0$. That is, the joint PMF can be expressed as the finite difference of the copula function, which includes 4 terms of copula distribution functions. For general p-th order Markov processes, the PMF involves 2^{p+1} terms of copula functions. Then the conditional PMF of Y_t given Y_{t-1} is given by

$$\begin{aligned}
&P(Y_t = y_t|Y_{t-1} = y_{t-1}, \underline{\boldsymbol{x}}_t)\\
=&\frac{P(Y_t = y_t, Y_{t-1} = y_{t-1}|\underline{\boldsymbol{x}}_t)}{P(Y_{t-1} = y_{t-1}|\underline{\boldsymbol{x}}_{t-1})}\\
=&\frac{C(u_t, u_{t-1};\boldsymbol{\alpha}) - C(u_t, v_{t-1};\boldsymbol{\alpha}) - C(u_{t-1}, v_t;\boldsymbol{\alpha}) + C(v_t, v_{t-1};\boldsymbol{\alpha})}{u_{t-1} - v_{t-1}}.
\end{aligned} \tag{2.3}$$

Thus, the likelihood function of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ can be derived as:

$$\begin{aligned} L(\boldsymbol{\beta},\boldsymbol{\alpha}) &= P(Y_1=y_1,Y_2=y_2,\ldots,Y_n=y_n|\underline{\boldsymbol{x}}_n;\boldsymbol{\beta},\boldsymbol{\alpha}) \\ &= P(Y_1=y_1|\boldsymbol{X}_1)P(Y_2=y_2|Y_1=y_1,\underline{\boldsymbol{x}}_2)\ldots P(Y_n=y_n|Y_{n-1}=y_{n-1},\underline{\boldsymbol{x}}_n) \\ &= \frac{\prod_{t=2}^{n}\{C(u_t,u_{t-1};\boldsymbol{\alpha})-C(u_t,v_{t-1};\boldsymbol{\alpha})-C(u_{t-1},v_t;\boldsymbol{\alpha})+C(v_t,v_{t-1};\boldsymbol{\alpha})\}}{\prod_{t=2}^{n-1}(u_t-v_t)}, \end{aligned}$$

and the log-likelihood function is given by

$$\begin{aligned} l(\boldsymbol{\beta},\boldsymbol{\alpha}) &= \log\{L(\boldsymbol{\beta},\boldsymbol{\alpha})\} \\ &= \sum_{t=2}^{n}\log\{C(u_t,u_{t-1};\boldsymbol{\alpha})-C(u_t,v_{t-1};\boldsymbol{\alpha})-C(u_{t-1},v_t;\boldsymbol{\alpha})+C(v_t,v_{t-1};\boldsymbol{\alpha})\} \\ &\quad -\sum_{t=2}^{n-1}\log(u_t-v_t), \end{aligned} \tag{2.4}$$

where $\boldsymbol{\beta}$ enters this expression through u and v. For scenarios with $p>1$, the likelihood takes a more complex form. A comprehensive derivation of the likelihood for general cases with $p \geq 1$ can be found in Section S1 of the Supplementary Material. The maximum likelihood estimator of $(\boldsymbol{\beta},\,\boldsymbol{\alpha})$ is obtained by

$$(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}}) = \underset{\boldsymbol{\beta},\boldsymbol{\alpha}}{\operatorname{argmax}}\, l(\boldsymbol{\beta},\boldsymbol{\alpha}).$$

For many copulas such as Gaussian and t copula, the copula function $C(\cdot)$ does not have a closed form, so numerical approximation is needed. If we use the finite difference likelihood (FDL) expression in (2.4), the estimator may be imprecise due to the accumulated approximation error involved in the four numerically computed orthant probabilities. The numerical error may even lead to negative values for the PMF in (2.2). The damage can be more severe when $p>1$, for which the joint PMF involes 2^{p+1} numerically computed orthant probabilities. Moreover, this FDL may also slow down the computation for larger p. To overcome this issue, we adapt the simulated likelihood (SL) proposed in Nikoloulopoulos

(2013) for Gaussian copula, which can be used to calculate the joint PMF with only one-time approximation:

$$\begin{aligned} l(\boldsymbol{\beta},\boldsymbol{\alpha}) &= \log\{P(Y_1 = y_1,\ldots,Y_n = y_n|\underline{\boldsymbol{x}}_n)\} \\ &= \log\{P(y_1 - 1 < Y_1 \le y_1,\ldots,y_n - 1 < Y_n \le y_n|\underline{\boldsymbol{x}}_t)\} \\ &= \sum_{t=2}^{n}\log\left(\int_{\Phi^{-1}(v_{t-1})}^{\Phi^{-1}(u_{t-1})}\int_{\Phi^{-1}(v_t)}^{\Phi^{-1}(u_t)}\phi_\alpha\left(z_1,z_2\right)dz_1dz_2\right) - \sum_{t=2}^{n-1}\log(u_t - v_t), \end{aligned} \tag{2.5}$$

where Φ denotes the standard normal CDF and ϕ_α denotes the bivariate normal density with correlation α for $p = 1$. The same technique can also be applied to other elliptical copulas such as t copula. For such copulas, since the method involves a one-time numerical approximation, it can improve the accuracy and reduce the computing time.

Throughout our implementation, we consider the simulated likelihood approach only for elliptical copulas such as Gaussian copula and t copula. For other copula families such as Archimedean copula with closed expressions for the copula distribution functions, the finite difference likelihood can be calculated directly using (2.4).

2.4 Point and interval prediction

For time series data analysis, one common question of interest is prediction at future time points. We now discuss the one-step-ahead prediction, that is, prediction of Y_{n+1}, based on the proposed model and estimator. The conditional distribution of Y_{n+1} given Y_n and $\underline{\boldsymbol{X}}_{n+1}$ can be well characterized by the conditional quantiles. Through conditional quantiles, we can not only construct point prediction of Y_{n+1} by using, e.g. the conditional median, but also prediction intervals. To accommodate discrete distributions, we adopt the generalized definition of quantile and define the qth quantile of a random variable Y as

$$Q_q(Y) = \inf\{y : P(Y \leq y) \geq q\},$$

for any given quantile level $0 < q < 1$.

For $\{Y_t, t = 1, \cdots, n\}$ satisfying Assumption A1, the conditional quantile function of Y_{n+1} given $(Y_n, \underline{\boldsymbol{X}}_{n+1})$ can be easily estimated. Define $G^{-1}(v|\boldsymbol{X}; \boldsymbol{\beta}) = \inf\{y : G(y|\boldsymbol{X}; \boldsymbol{\beta}) \geq v\}$. Note that $G^{-1}(v|\boldsymbol{X}; \boldsymbol{\beta})$ is a monotonic transformation of v. Hence the q-th conditional quantile of Y_{n+1} is given by

$$\begin{aligned} Q_q\left(Y_{n+1}|Y_1, \ldots, Y_n, \underline{\boldsymbol{X}}_{n+1}\right) &= Q_q\left(Y_{n+1}|Y_n, \underline{\boldsymbol{X}}_{n+1}\right) \\ &= G^{-1}\left[Q_q\left\{U_{n+1}|U_n = G\left(Y_n|\underline{\boldsymbol{X}}_{n+1}; \boldsymbol{\beta}\right); \boldsymbol{\alpha}\right\} \left|\underline{\boldsymbol{X}}_{n+1}; \boldsymbol{\beta}\right.\right], \quad (2.6) \end{aligned}$$

where $Q_q\left(U_{n+1}|U_n; \boldsymbol{\alpha}\right)$ is the conditional quantile function of U_{n+1} given U_n, that is,

$$Q_q\left(U_{n+1}|U_n = u; \boldsymbol{\alpha}\right) = C_{2|1}^{-1}\left(q|u; \boldsymbol{\alpha}\right),$$

in which $C_{2|1}\left(\cdot|u; \boldsymbol{\alpha}\right) = \partial C\left(u, \cdot; \boldsymbol{\alpha}\right)/\partial u$ is the conditional distribution of U_{n+1} given $U_n = u$. It follows from (2.4) and (2.5) that the plug-in estimator of the conditional quantile of U_{n+1} given $U_n = u$ is

$$\widehat{Q}_q^U(u) = Q_q^U(u; \hat{\boldsymbol{\alpha}}) = C_{2|1}^{-1}\left(q|u; \hat{\boldsymbol{\alpha}}\right),$$

and the plug-in estimator of the qth conditional quantile Y_{n+1} given $(Y_n = y, \underline{\boldsymbol{X}}_n = \underline{\boldsymbol{x}}_n)$ is

$$\begin{aligned} \widehat{Q}_q^Y(Y|y, \underline{\boldsymbol{x}}_{n+1}) &= G^{-1}\left[\widehat{Q}_q^U\left\{G\left(y|\underline{\boldsymbol{x}}_{n+1}; \hat{\boldsymbol{\beta}}\right); \hat{\boldsymbol{\alpha}}\right\}\left|\underline{\boldsymbol{x}}_{n+1}; \hat{\boldsymbol{\beta}}\right.\right] \\ &= G^{-1}\left[C_{2|1}^{-1}\left\{q|G\left(y|\underline{\boldsymbol{x}}_{n+1}; \hat{\boldsymbol{\beta}}\right); \hat{\boldsymbol{\alpha}}\right\}\left|\underline{\boldsymbol{x}}_{n+1}; \hat{\boldsymbol{\beta}}\right.\right]. \end{aligned}$$

After the conditional quantiles are estimated, we can construct the $100(1-\nu)\%$ prediction interval $[L, U]$ for the one-step ahead observation Y_{n+1}, which satisfies

$$P(L \leq Y_{n+1} \leq U|\boldsymbol{X}; \hat{\boldsymbol{\beta}}) \geq 1 - \nu. \quad (2.7)$$

For continuous distributions, a standard method to construct prediction intervals is to use appropriate pivotal quantities. This is not possible for discrete distributions since such pivotal quantities are not available; see Bain and Patel (1993) and Kim et al. (2021) for some related discussions. If we simply take $L = \widehat{Q}^Y_{\nu/2}(Y|y, \underline{\boldsymbol{x}}_{n+1})$ and $U = \widehat{Q}^Y_{1-\nu/2}(Y|y, \underline{\boldsymbol{x}}_{n+1})$, we have $P(Y \leq L; \boldsymbol{\theta}|\boldsymbol{X}) \geq \nu/2$ and $P(Y \geq U; \boldsymbol{\theta}|\boldsymbol{X}) \geq 1-\nu/2$, and (2.7) is not ensured to hold. To ensure the coverage probability of the prediction interval, we define the lower and upper bound of the prediction interval as follows:

$$L = \sup\{y : G(y-1|\boldsymbol{X}; \hat{\boldsymbol{\beta}}) \leq \nu/2\},\ U = \inf\{y : G(y|\boldsymbol{X}; \hat{\boldsymbol{\beta}}) \geq 1-\nu/2\}. \tag{2.8}$$

By doing so, we can ensure that (2.7) holds.

2.5 Asymptotic properties

Denote $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$ as the parameter vector with dimension m, $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T)^T$ as the true value, and $\boldsymbol{\Theta}$ as the parameter space. Let $l_n(\boldsymbol{\theta})$ and $s_n(\boldsymbol{\theta})$ denote the log-likelihood and its first order derivatives, respectively. Define $f_t(y_t; \boldsymbol{\theta}) = P(Y_t = y_t|Y_{t-1} = y_{t-1}, \underline{\boldsymbol{x}}_t; \boldsymbol{\theta})$. The following set of conditions are needed to ensure the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$.

A2. (a) $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^m$, $m \in \mathbb{N}$; (b) $f_t(y_t; \boldsymbol{\theta})$ is continuous on $\boldsymbol{\Theta}$ a.s.; (c) $E[l_n(\boldsymbol{\theta})]$ has identifiably unique maximizer interior to $\boldsymbol{\Theta}$.

A3. $E[\log\{f_t(y_t; \boldsymbol{\theta})\}]$ exists, and it is finite and continuous on $\boldsymbol{\Theta}$.

A4. Given $\underline{\boldsymbol{X}}_t$, the process $\{Y_t, t = 1, \ldots, n\}$ is β-mixing with $\beta = O(t^{-b})$ for $b > \frac{r-1}{r}$, $r > 1$.

A5. $f_t(y_t; \boldsymbol{\theta})$ is continuously differentiable of order 2.

A6. (a) $n^{-1}\nabla E[l_n(\boldsymbol{\theta})] = n^{-1}E[\nabla l_n(\boldsymbol{\theta})] < \infty$; (b) $n^{-1}\nabla^2 E[l_n(\boldsymbol{\theta})] = n^{-1}E[\nabla^2 l_n(\boldsymbol{\theta})] < \infty$; (c) $E[\nabla^2 f_t(y_t; \boldsymbol{\theta})]$ is continuous on $\boldsymbol{\Theta}$.

A7. Let $A_{n0} = n^{-1}\nabla^2 E[l_n(\boldsymbol{\theta}_0)]$, and A_{n0} is $O(1)$ and negative definite uniformly in n.

A8. (a) $E[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \log\{f_t(y_t;\boldsymbol{\theta})\}^2] < \infty$; (b) $E[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\nabla^2 \log\{f_t(y_t;\boldsymbol{\theta})\}|^2] < \infty$; (c) $E[\{\nabla f_t(y_t;\boldsymbol{\theta})\}^2] < \infty$.

Theorem 1. *Suppose that A1 - A8 hold, then we have (i)* $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$*; (ii)* $B_{n0}^{-1/2} A_{n0}\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(0, I_m)$*, where* $A_{n0} = n^{-1}\nabla^2 E[l_n(\boldsymbol{\theta}_0)]$ *and* $B_{n0} = n^{-1}Var(s_n(\boldsymbol{\theta}_0))$*.*

Assumptions A2, A3, A5 and A6 can be easily verified. Assumption A4 concerns the mixing property of the process. Beare (2012, 2010) have shown that series generated from most Archimedean copulas satisfy A4, while Chen and Fan (2006) and Chen et al. (2009) established the mixing properties for Gaussian and t copulas, respectively, in cases where the marginal distributions are continuous. These results, along with Lemma 1 presented in the supplementary material, ensure that A4 holds for most Archimedean, Gaussian, and t copulas. Assumptions A7 and A8 are similar to A1(ii) and A5 in Chen and Fan (2006).

2.6 Copula selection

In Section 2.2, we discussed the identifiability issue and presented one sufficient condition to ensure the identifiability of copula under the proposed model. Resolving the identifiability issue allows us to select a parametric form for the copula from existing copula families.

Focusing on multivariate discrete responses, Yang et al. (2020) suggested a method to select copula by minimizing a L_2-norm distance between the fitted parametric copulas with their proposed nonparametric copula estimator. We considered this approach in our setup, but our numerical studies show that this method is time consuming with unstable performance. The main challenge is that the criterion is based on the integration of squared distance, which requires repeated kernel-type estimation at multiple points. The computation and performance are further complicated by the selection of bandwidth involved in the smoothing. To overcome these challenges, we propose a cross-validation (CV) procedure to select copula for count time series data. Let k be a pre-set minimum number of observations in the training data. We repeat the validation sequentially for a total of $n-k$ times. To assess the performance of the predictive models for count data based on a candidate copula, we focus on probabilistic forecasts, that is, evaluating the predictive distribution generated by the copula and parameters estimated by the training data. The reason we focus on probabilistic forecasts instead of using the commonly used mean squared error (MSE) for point prediction is that count data is often skewed so that the MSE will be dominated by predictions at the tails. The detailed procedure is as follows.

For the ith cross-validation, choose observations at time $1, 2, \ldots, k+i-1$ as the training data and let the observation at time $k+i$ be the test data. Fit the proposed model with the candidate copula by using the training data and let $\hat{\boldsymbol{\theta}}^*$ denote the estimated parameter. The predictive distribution for Y_t at $t = k+i$ can be obtained by $P(Y_t = y|Y_{t-1}, \underline{\boldsymbol{X}}_t; \hat{\boldsymbol{\theta}}^*) \doteq \hat{P}^*$; see the expression in (2.3). To assess the performance of the resulting predictive distribution, we adapt the logarithmic score (LOGS), proposed in (Czado et al., 2009), at time $t = k+i$, defined as

$$\text{LOGS}(\hat{P}^*, y_t) = -\log\{P(Y_t = y_t|Y_{t-1} = y_{t-1}, \underline{\boldsymbol{X}}_t); \hat{\boldsymbol{\theta}}^*\},$$

where y_t is the observed response at time t. LOGS is a measure that evaluates probabilistic forecasts of count data based on a predictive distribution and observations at testing time. A smaller LOGS value indicates that the predictive distribution fits the data more closely.

We repeat the above step for $i = 1, 2, \ldots, n-k$, and compute the average of LOGS across $n-k$ validations. The copula is then chosen to be the one that gives the smallest average LOGS. The CV procedure is also applicable to determine the Markovian order p and this process can be conducted simultaneously with the selection of the copula.

3. Simulation study

We conduct a simulation to compare the finite sample performance of the proposed method with existing approaches for analyzing count time series data. In Sections 3.1-3.2, we consider generating data $\{Y_t, t = 1, \ldots, n\}$ from the Markov process with bivariate copula $C(u, v; \alpha)$ and Poisson regression model $Y_t|\underline{\boldsymbol{X}}_t \sim Poisson(\lambda_t)$, where $\log(\lambda_t) = \beta_1 + \beta_2 X_t$. Below are the steps used for data generation. Specifically, we first generate an *i.i.d.* sequence from Unif$(0, 1)$, denoted as $\{v_t\}_{t=1}^n$, and $\{X_t, t = 1, \ldots, n\}$ from $N(0, 1)$. Let $U_1 = v_1$ and $U_t = C_{2|1}^{-1}(v_t|U = U_{t-1}; \alpha)$ for $t = 2, \ldots, n$. Then for $t = 1, \ldots, n$, we set $Y_t = \inf\{y : G(y; \lambda_t) \geq U_t\}$, where $G(y; \lambda_t) = P(Y_t \leq y|\underline{\boldsymbol{X}}_t)$ is the CDF of Poisson(λ_t) with $\log(\lambda_t) = \beta_1 + \beta_2 X_t$. To reach the stationarity of U_t, we generate sequences with a length of burn-in periods of 2,000. We consider two sample sizes, $n = 200$ and 500, and two different levels of temporal dependence with Kendall's $\tau = 0.3$ and 0.7 for Joe, Gaussian, and t copulas, respectively. Kendall's τ is a measure of rank correlation between two variables, defined as the probability of concordance minus the probability of discordance (Kendall, 1938; Nelsen, 2006). We consider two cases for data generating process: Case 1 with $\beta_1 = 2$, $\beta_2 = 1$, and Case 2 with $\beta_1 = 0$, $\beta_2 = 1$. In Section 3.3, we consider an additional simulation design where data are generated from the Poisson Autoregressive (PA) model: $Y_t|(Y_{t-1}, X_t) \sim Poisson(\exp(a + b\log(Y_{t-1} + 1) + cX_t))$ for $t = 2, \ldots, n$ with $a = 0.8$, $b = 0.4$, $c = 0.5$ and $\{X_t, t = 1, \ldots, n\} \sim^{i.i.d} N(0, 1)$. For each scenario, the simulation is repeated 500 times. Henceforth, we will refer to the proposed estimation method for the Copula-based Markov

Model for Count data as "COMC."

3.1 Parameter estimation

We first assess the performance of the proposed estimator for the copula parameter α and the marginal parameters (β_1, β_2) assuming the correct specification of the copula function. For data generated with Gaussian copula, the model assumption in Jia et al. (2021) also holds and parameters in the two models coincide. Therefore, besides the proposed COMC estimator, we also include the Gaussian copula-based particle filtering estimator (PF) in Jia et al. (2021) as a comparison. In the proposed method, we use the simulated likelihood (SL) method to calculate the likelihood for data generated with a Gaussian copula.

Table 1 summarizes the averaged bias and root mean squared error (RMSE) of COMC and PF estimators when $\alpha = 0.454$ in Case 1 and 0.891 in Case 2, corresponding to Kendall's $\tau = 0.3$ and 0.7. Figure 1 shows boxplots of estimations across 500 simulations. The white plots correspond to $n = 200$ while the grey ones correspond to $n = 500$. The first row is for Case 1 and the second row is for Case 2. The dashed lines correspond to the true parameter values. Results suggest that the two methods are comparable in Case 1 with a weaker temporal dependence. However, for Case 2 with a higher temporal dependence, COMC gives significantly more efficient estimations than PF.

3.2 Conditional quantile estimation with selected copula

We assess the performance of the proposed estimator of $Q_q^Y(Y_{n+1}|Y_n, X)$, the q-th conditional quantile of Y_{n+1} at time $n+1$ given the observed data, when the true copula is unknown and chosen from a class of candidate families, including Gaussian, t, Joe, Frank and Clayton copulas. Note that here we only consider these 5 candidate families, and in practice, any

Table 1: 1000×bias and 1000×RMSE of different estimators of β_1, β_2 and α in Cases 1-2 with Gaussian copula.

			$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\alpha}$	
Case	n	Method	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	200	COMC	-1.2 (2.0)	43.7 (1.5)	0.6 (1.0)	21.6 (0.9)	-1.5 (2.4)	54.7 (1.6)
		PF	-1.4 (2.0)	44.3 (1.5)	0.5 (1.0)	21.3 (0.9)	-0.0 (2.5)	55.3 (1.6)
	500	COMC	-0.1 (1.2)	27.7 (0.8)	0.7 (0.6)	13.3 (0.4)	-2.3 (1.6)	35.8 (1.1)
		PF	-0.4 (1.3)	28.9 (0.9)	0.9 (0.6)	13.7 (0.5)	0.9 (1.7)	37.2 (1.1)
2	200	COMC	7.4 (10.2)	227.0 (7.4)	1.4 (3.3)	73.5 (2.4)	2.6 (1.0)	23.0 (0.7)
		PF	-39.5 (12.7)	286.6 (10.3)	12.2 (4.3)	96.2 (3.6)	-19.0 (1.2)	32.4 (1.1)
	500	COMC	0.9 (6.7)	150.2 (4.7)	-0.2 (2.2)	48.5 (1.5)	-0.4 (0.6)	14.0 (0.5)
		PF	-18.5 (9.1)	205.1 (6.9)	6.4 (3.0)	68.1 (2.3)	-34.5 (1.0)	40.7 (1.0)

COMC: copula-based Morkov model for count data; PF: particle filtering. Values in the parentheses are standard errors.

copula family can be considered. We consider Cases 1-2 and two quantile levels $q = 0.5$ and 0.9. For comparison, we include two estimators from COMC: COMC_{TC}, COMC_{SC}, and two other methods: PF and PA. The method COMC_{TC} provides estimation with the true copula, while COMC_{SC} gives estimation with the selected copula. The method PF is the Gaussian copula-based particle filtering approach proposed in Jia et al. (2021). The PA method is the Poisson autoregression method proposed in Fokianos and Tjøstheim (2011).

Table 2 summarizes the copula selection results based on the cross-validation method. Results show that the cross-validation method works well for selecting the correct copula, and the selection accuracy increases with the sample size. Table 3 summarizes the MSE of

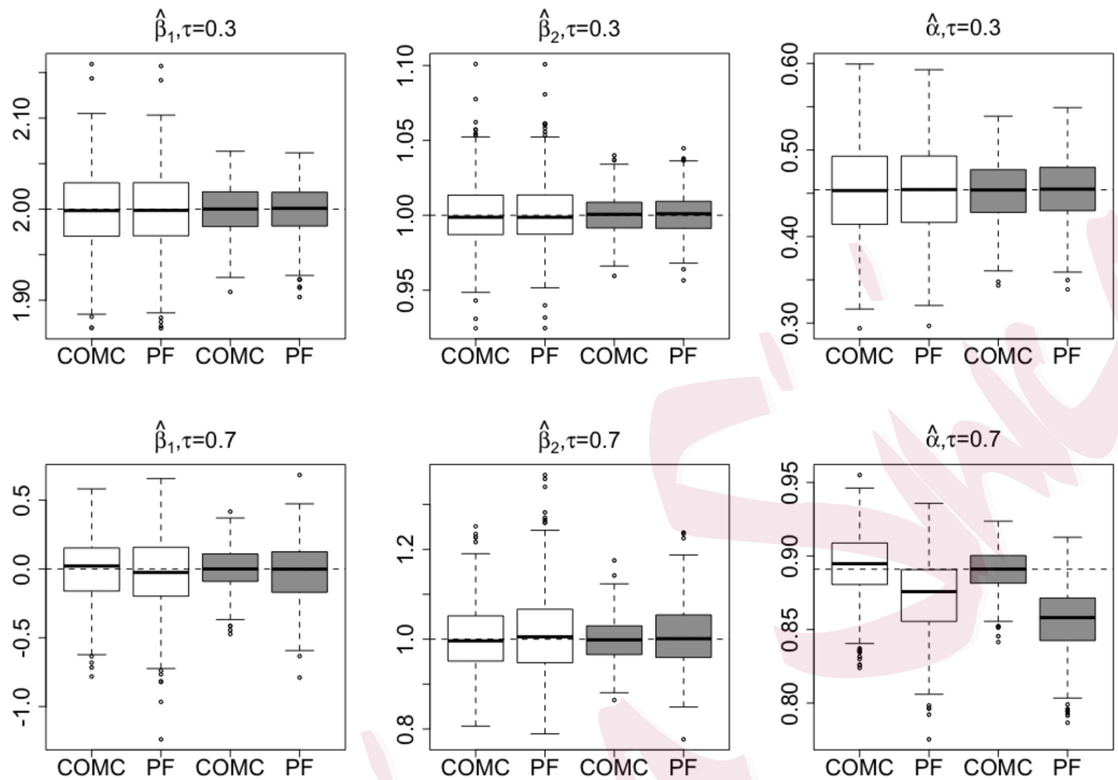

Figure 1: Boxplots of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\alpha}$ for $n = 200$ (white) and 500 (grey) across 500 simulations with α corresponding to $\tau = 0.3$ and $\tau = 0.7$. True values are shown as a horizontal dashed line. COMC: copula-based Morkov model for count data; PF: particle filtering.

$\widehat{Q}_q(Y_{n+1}|Y_n, X)$ from four methods. As expected, COMC_{TC} is the most efficient estimator as it is based on the correct copula function. In all cases considered, the COMC_{SC} estimator is competitive to COMC_{TC}, and it has smaller MSE than PF and PA, even when the copula was occasionally misspecified in COMC_{SC}.

3.3 Prediction at future time points

We consider the following metrics to assess the prediction performance of different methods at time $t = n + 1$: (i) the coverage percentages (CPs) and average lengths (ALs) of the

Table 2: Percentages of copulas chosen by cross-validation across 500 simulations.

			Selected copula				
Case	True	n	Clayton	Gumbel	Joe	Gaussian	t
1	Joe	200	0.2	21.0	74.6	1.6	2.6
		500	0.0	15.6	83.8	0.2	0.4
	Gaussian	200	5.6	17.0	3.6	52.0	21.8
		500	0.2	10.2	0.0	67.2	22.4
	t_2	200	4.4	29.0	6.4	6.0	54.2
		500	0.8	17.6	1.2	1.0	79.4
2	Joe	200	0.2	19.0	77.6	1.0	2.2
		500	0.0	9.6	90.4	0.0	0.0
	Gaussian	200	12.2	15.8	3.6	48.2	20.2
		500	3.0	8.0	0.0	72.8	16.2
	t_2	200	3.6	8.6	4.0	1.0	82.8
		500	0.8	0.2	0.0	0.0	99.0

90% prediction intervals for predicting Y_{n+1}; (ii) logarithmic score (LOGS) of the predictive conditional distribution P and the realized count Y_{n+1}. For comparison, we include three methods from COMC: COMC_0, COMC_{TC}, COMC_{SC} and two other methods: PF and PA. To evaluate the sensitivity of various methods to model misspecification, we generated data using three different copulas (Cases 1 and 2) within the copula-based Markov model and also from the Poisson Autoregressive (PA) model. The method COMC_0 is based on the true copula and true parameters, while COMC_{TC} provides estimations using the true copula.

In contrast, COMCSC uses the selected copula for estimation. Note that both COMC_0 and COMCTC are not applicable to the PA model. Due to the discrete nature of the data, achieving an exact 90% prediction interval coverage is infeasible. Therefore, we use the coverage percentage of method COMC_0 as the baseline in Cases 1 and 2, and the PA method as the baseline in the PA model when comparing methods. Table 4 summarizes the coverage percentages (CPs) and the average lengths (ALs) of 90% prediction intervals, and Table 5 summarizes the logarithmic score (LOGS) from the five methods. In Cases 1 and 2, where data were generated from the copula-based Markov model, the proposed COMC_{SC} performs comparably to COMC_{TC} and COMC_0. When the true copula is Gaussian, PF and COMC_{SC} exhibit similar performance. However, COMC_{SC} outperforms PF and PA for all other copulas: it provides prediction intervals with coverage similar to that of COMC_0 but shorter than those of PF and PA. The smaller LOGS values obtained from COMC_{SC} also suggest higher prediction accuracy compared to PF and PA. When data were generated from the PA model, for which the model assumed by COMC is misspecified, COMC_{SC} demonstrates performance similar to PF, albeit slightly worse than the PA method in terms of LOGS; however, this difference is not statistically significant.

4. Real data analysis

In this section, we analyze the count time series of daily deaths from COVID-19 in Virginia. The dataset contains the daily deaths in Virginia due to COVID-19 from June 22, 2020, to December 31, 2021, which is maintained and archived by the Centers for Disease Control and Prevention. The dataset contains 400 daily observations in Virginia, with the first 320 time points being used as training data, while the last 80 being used as testing data. We include linear and quadratic terms of time and the number of confirmed cases 10 days ago as covariates.

The data are over-dispersed, with a sample mean of 29.3 and a sample variance of 1063.9. We choose the marginal model to be generalized Poisson (GP) with the dispersion parameter $\varphi > 1$ fixed in time and the rate parameter λ_t varying in time with the log link:

$$\log(\lambda_t) = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 \log(x_{t-10} + 1),$$

where x_{t-10} is the number of confirmed cases 10 days ago. In this application, t, t^2 and x_{t-10} can all be considered as continuous covariates due to their large ranges. We conduct a cross-validation with $k = 320$ to choose the order p and copula among five families: Gaussian, t, Joe, Gumbel, and Clayton. Based on the cross-validation, we choose the model with t_2 copula and $p = 1$ corresponding to the first-order Markov process.

By fitting the model using the full sample, we obtain the coefficient estimates and standard errors (in the parentheses) as follows: $\hat{\alpha} = 0.82(0.02)$, $\hat{\beta}_1 = -0.88(0.29)$, $\hat{\beta}_2 = -0.0016(0.0011)$, $\hat{\beta}_3 = 3.1 \times 10^{-6}(2.6 \times 10^{-6})$, $\hat{\beta}_4 = 0.60(0.04)$ and $\hat{\varphi} = 7.64(0.07)$. Kendall's τ corresponds to a t copula with a coefficient of 0.82 is 0.61, indicating a strong first-order temporal dependence. Marginal model estimates confirm the overdispersion of data and suggest that the mean death rate at day t, λ_t, relates positively to x_{t-10}.

We conduct a cross-validation using the last 80 observations as the testing data to compare the prediction accuracy of the proposed method with four alternatives: (1) the integer-valued ARCH model (INARCH) with log link, which assumes $Y_t|(Y_{t-1}, X_t) \sim GP(\exp(a + b_1 \log(Y_{t-1} + 1) + c_1 t + c_2 t^2 + c_3 \log(X_{t-10} + 1)), \varphi)$; (2) the integer-valued GARCH model (INGARCH) with log link, which assumes $Y_t|(Y_{t-1}, X_t) \sim GP(\exp(a + b_1 \log(Y_{t-1} + 1) + b_2 \log(\lambda_{t-1}) + c_1 t + c_2 t^2 + c_3 \log(X_{t-10} + 1)), \varphi)$; (3) the INGARCH model with softplus link

(Softplus-INGARCH), which assumes $Y_t|(Y_{t-1}, X_t) \sim GP(\log(1+\exp(a+b_1\log(Y_{t-1}+1)+b_2\log(\lambda_{t-1})+c_1t+c_2t^2+c_3\log(X_{t-10}+1))), \varphi)$, and (4) the particle filtering (PF) method of Jia et al. (2021) assuming the latent process follows an AR(1) model. For more details on these alternative models, refer to Cui and Zhu (2018), Cui et al. (2020), Fokianos et al. (2020) and Weiß et al. (2022). Table 6 summarizes the coverage percentages and average lengths of 95% prediction intervals, the LOGS and mean squared relative prediction error (MSRPE) of point predictions from various methods for one-step-ahead predictions, along with the AIC and BIC values. Results suggest that prediction intervals from log-INARCH, log-INGARCH and Softplus-INGARCH all exhibit notably low coverage. While both COMC and PF provide prediction intervals with similar coverage, those from PF are significantly wider. In terms of point predictions, COMC also demonstrate advantages over all the competitors by yielding lower LOGS and MSPRE values. Additionally, the COMC method yields the smallest AIC and BIC values, indicating a better fit to the data.

Supplementary material

The online supplementary material includes technical proofs, a discussion of the basic statistics of the proposed model, and the derivation of the likelihood for $p > 1$.

Acknowledgements

This work was partly supported by an IR/D program and grant DMS-1712760 from the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Table 3: 100×MSE of $\widehat{Q}_q(Y_{n+1}|Y_n)$ from different methods at $q = 0.5$ and $q = 0.9$.

Model	Case	n	Quantile	COMC_{TC}	COMC_{SC}	PF	PA
Joe	1	200	0.5	6.0 (1.1)	6.6 (1.3)	15.6 (1.9)	252.8 (29.7)
			0.9	5.8 (1.0)	7.2 (1.2)	28.7 (2.5)	200.8 (37.5)
	1	500	0.5	2.6 (0.7)	2.6 (0.7)	20.0 (3.1)	291.0 (34.2)
			0.9	4.4 (1.2)	4.4 (1.2)	38.6 (4.9)	231.8 (36.8)
	2	200	0.5	27.6 (3.4)	40.4 (8.1)	181.2 (28.7)	749.4 (108.8)
			0.9	26.8 (4.2)	28.2 (4.3)	108.8 (11.0)	486.6 (51.4)
	2	500	0.5	19.0 (2.6)	20.4 (2.8)	159.6 (28.5)	605.0 (89.2)
			0.9	16.8 (2.5)	19.4 (2.6)	106.2 (14.8)	444.4 (49.8)
t_2	1	200	0.5	2.4 (0.7)	4.0 (0.9)	17.8 (2.6)	222.0 (18.3)
			0.9	6.4 (1.1)	10.4 (1.5)	27.2 (3.0)	157.8 (20.8)
	1	500	0.5	1.2 (0.5)	2.0 (0.6)	18.0 (2.1)	241.6 (23.0)
			0.9	4.4 (0.9)	6.8 (1.1)	29.6 (2.3)	186.0 (23.7)
	2	200	0.5	20.4 (7.5)	50.2 (18.1)	146.2 (73.1)	752.2 (324.4)
			0.9	29.4 (16.3)	54.4 (20.5)	612.2 (443.3)	392.0 (118.1)
	2	500	0.5	8.6 (1.3)	9.0 (1.5)	73.4 (18.9)	376.0 (65.3)
			0.9	11.2 (2.2)	11.8 (2.2)	118.2 (11.7)	298.2 (42.1)
Gaussian	1	200	0.5	8.0 (2.1)	7.2 (1.3)	11.0 (2.2)	246.8 (31.3)
			0.9	15.8 (3.6)	16.4 (2.0)	15.0 (2.5)	233.8 (34.5)
	1	500	0.5	5.2 (1.0)	4.8 (1.0)	7.6 (1.2)	183.2 (14.6)
			0.9	7.4 (1.2)	8.0 (1.2)	11.2 (1.4)	170.0 (18.2)
	2	200	0.5	69.8 (25.8)	70.4 (25.9)	65.8 (21.8)	400.0 (74.1)
			0.9	81.0 (26.0)	85.0 (26.1)	73.8 (21.9)	453.8 (90.2)
	2	500	0.5	70.8 (51.2)	74.2 (51.3)	79.6 (57.8)	350.0 (107.2)
			0.9	71.4 (51.2)	75.2 (51.3)	75.2 (51.2)	423.8 (146.9)

COMC_{TC}: COMC based on true copula; COMC_{SC}: COMC based on selected copula; PF: particle filtering method from Jia et al. (2021); PA: Poisson autoregression method from Fokianos and Tjøstheim (2011). Values in the parentheses are standard errors.

Table 4: Coverage percentages and 10×average lengths of 90% prediction intervals for predicting Y_{n+1}.

Model	Case	n	Coverage Percentages					10×Average Lengths				
			$COMC_0$	$COMC_{TC}$	$COMC_{SC}$	PF	PA	$COMC_0$	$COMC_{TC}$	$COMC_{SC}$	PF	PA
Joe	1	200	97.4	97.2	97.0	98.2	98.4	18.4	18.3	18.7	24.1	34.2
		500	96.2	96.0	96.2	97.6	99.0	19.5	19.8	20.1	26.8	37.6
	2	200	96.4	96.4	96.2	96.6	93.8	93.1	93.0	93.4	97.7	107.4
		500	95.0	95.2	95.4	96.0	93.8	92.0	91.8	91.9	96.4	106.6
t_2	1	200	98.2	97.8	97.4	97.8	98.8	23.4	23.7	23.3	26.1	34.2
		500	97.0	96.2	96.6	97.4	98.8	23.2	23.3	23.1	27.3	34.6
	2	200	94.4	94.4	94.0	94.4	96.6	95.5	94.8	94.4	96.7	106.2
		500	96.4	96.6	96.6	96.2	96.2	93.0	93.2	93.2	98.2	106.7
Gaussian	1	200	98.0	98.2	97.8	98.4	99.0	24.8	24.7	24.6	26.1	36.1
		500	99.0	99.0	99.0	99.2	98.0	24.1	24.2	23.9	25.9	34.8
	2	200	95.8	96.2	96.8	96.2	95.8	99.3	99.4	99.6	99.1	108.6
		500	95.0	94.4	94.2	94.0	93.0	100.7	100.2	100.3	99.9	109.9
PA		200			92.0	93.4	95.0			6.8	6.8	6.9
		500			96.0	96.2	97.0			6.6	6.5	6.7

$COMC_0$: COMC based on the true copula and true parameters; $COMC_{TC}$: COMC based on the true copula; $COMC_{SC}$: COMC based on the selected copula; PF: particle filtering method from Jia et al. (2021); PA: Poisson autoregression method from Fokianos and Tjøstheim (2011). Values in the parentheses are standard errors.

Table 5: 100× LOGS of prediction for Y_{n+1} by different methods.

Model	Case	n	COMC$_0$	COMC$_{\text{TC}}$	COMC$_{\text{SC}}$	PF	PA
Joe	1	200	80.4 (3.6)	79.8 (3.6)	79.7 (3.5)	85.2 (3.4)	125.0 (3.6)
		500	84.0 (3.4)	85.0 (3.5)	85.1 (3.5)	93.0 (3.3)	133.3 (3.5)
	2	200	221.6 (4.0)	221.0 (3.9)	221.6 (3.9)	229.8 (4.0)	243.0 (4.1)
		500	225.8 (3.9)	226.1 (3.9)	225.8 (3.9)	229.6 (3.9)	243.7 (4.0)
t_2	1	200	89.5 (3.6)	89.8 (3.7)	90.0 (3.7)	93.7 (3.7)	124.5 (3.4)
		500	91.4 (4.4)	92.1 (4.5)	93.8 (5.0)	100.1 (5.2)	129.2 (3.7)
	2	200	223.4 (4.2)	223.6 (4.2)	225.0 (4.3)	233.5 (5.7)	239.6 (4.1)
		500	217.0 (3.7)	217.0 (3.7)	217.1 (3.6)	222.9 (4.0)	233.1 (3.6)
Gaussian	1	200	90.8 (3.5)	92.6 (3.7)	93.1 (3.7)	93.0 (3.5)	127.6 (3.6)
		500	87.6 (3.6)	87.3 (3.5)	87.3 (3.6)	87.9 (3.3)	125.2 (3.7)
	2	200	230.4 (3.9)	231.0 (3.9)	230.3 (3.9)	231.1 (3.9)	239.7 (3.8)
		500	232.9 (3.9)	233.5 (4.0)	233.7 (4.0)	233.6 (4.0)	246.8 (4.3)
PA		200			224.8 (4.5)	223.7 (4.4)	217.2 (3.8)
		500			211.8 (4.6)	212.1 (4.9)	207.7 (3.8)

COMC$_0$: COMC based on the true copula and true parameters; COMC$_{\text{TC}}$: COMC based on the true copula; COMC$_{\text{SC}}$: COMC based on the selected copula; PF: particle filtering method from Jia et al. (2021); PA: Poisson autoregression method from Fokianos and Tjøstheim (2011). Values in the parentheses are standard errors.

Table 6: Performance of different methods for one-step-ahead prediction of COVID-19 data.

Method	95% prediction interval CP	95% prediction interval AL	LOGS	MSRPE	AIC	BIC
COMC	97.5	89.3 (2.5)	4.12 (0.07)	0.14 (0.07)	3206	3210
PF	98.8	107.7 (3.3)	4.25 (0.06)	0.17 (0.03)	3277	3281
Log-INARCH	70.0	23.5 (0.4)	5.17 (0.49)	0.19 (0.07)	5657	5677
Log-INGARCH	74.0	23.4 (0.4)	5.49 (0.55)	0.19 (0.07)	5405	5729
Softplus-INGARCH	60.0	19.8 (0.3)	5.52 (0.48)	0.16 (0.04)	5311	5312

COMC: the proposed copula-based Morkov model for count data; PF: particle filtering method; log-INARCH: integer-valued ARCH model with the log link; log-INGARCH: integer-valued GARCH model with the log link; Softplus-INGARCH: INGARCH model with the softplus link; CP: coverage percentage; AL: average length; MSRPE: mean squared relative prediction error. Values in the parentheses are standard errors.

References

Bain, L. J. and J. K. Patel (1993). Prediction intervals based on partial observations for some discrete distributions. *IEEE Transactions on Reliability 42*(3), 459–463.

Beare, B. K. (2010). Copulas and temporal dependence. *Econometrica 78*(1), 395–410.

Beare, B. K. (2012). Archimedean copulas and temporal dependence. *Econometric Theory*, 1165–1185.

Brandt, P. T., J. T. Williams, B. O. Fordham, and B. Pollins (2000). Dynamic modeling for persistent event-count time series. *American Journal of Political Science*, 823–843.

Chen, X. and Y. Fan (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics 130*(2), 307–335.

Chen, X., W. B. Wu, Y. Yi, et al. (2009). Efficient estimation of copula-based semiparametric Markov models. *The Annals of Statistics 37*(6B), 4214–4253.

Cui, Y., Q. Li, and F. Zhu (2020). Flexible bivariate poisson integer-valued garch model. *Annals of the Institute of Statistical Mathematics 72*(6), 1449–1477.

Cui, Y. and F. Zhu (2018). A new bivariate integer-valued garch model allowing for negative cross-correlation. *Test 27*, 428–452.

Czado, C., T. Gneiting, and L. Held (2009). Predictive model assessment for count data. *Biometrics 65*(4), 1254–1261.

Davis, R. A., W. T. Dunsmuir, and S. B. Streett (2003). Observation-driven models for poisson counts. *Biometrika 90*(4), 777–790.

Davis, R. A., W. T. Dunsmuir, and Y. Wang (2000). On autocorrelation in a Poisson regression model. *Biometrika 87*(3), 491–505.

Davis, R. A. and H. Liu (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, 1673–1707.

Davis, R. A. and R. Wu (2009). A negative Binomial model for time series of counts. *Biometrika 96*(3), 735–749.

Denuit, M. and P. Lambert (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis 93*(1), 40–57.

Elsner, J. B. and T. H. Jagger (2006). Prediction models for annual us hurricane counts. *Journal of Climate 19*(12), 2935–2952.

Ferland, R., A. Latour, and D. Oraichi (2006). Integer-valued garch process. *Journal of Time Series Analysis 27*(6), 923–942.

Fokianos, K., A. Rahbek, and D. Tjøstheim (2009). Poisson autoregression. *Journal of the American Statistical Association 104*(488), 1430–1439.

Fokianos, K., B. Støve, D. Tjøstheim, P. Doukhan, et al. (2020). Multivariate count autoregression. *Bernoulli 26*(1), 471–499.

Fokianos, K. and D. Tjøstheim (2011). Log-linear poisson autoregression. *Journal of Multivariate Analysis 102*(3), 563–578.

Genest, C. and J. Nešlehová (2007). A primer on copulas for count data. *ASTIN Bulletin: The Journal of the IAA 37*(2), 475–515.

Heinen, A. and E. Rengifo (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance 14*(4), 564–583.

Hughes, J. (2015). copCAR: A flexible regression model for areal data. *Journal of Computational and Graphical Statistics 24*(3), 733–755.

Jia, Y., S. Kechagias, J. Livsey, R. Lund, and V. Pipiras (2021). Latent Gaussian count time series. *Journal of the American Statistical Association*, 1–28.

Jung, R. C., M. Kukuk, and R. Liesenfeld (2006). Time series of count data: modeling, estimation and diagnostics. *Computational Statistics & Data Analysis 51*(4), 2350–2364.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika 30*(1/2), 81–93.

Kim, T., B. Lieberman, G. Luta, and E. A. Peña (2021). Prediction intervals for poisson-based regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, e1568.

Livsey, J., R. Lund, S. Kechagias, and V. Pipiras (2018). Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *The Annals of Applied Statistics 12*(1), 408–431.

Madsen, L. (2009). Maximum likelihood estimation of regression parameters with spatially dependent discrete data. *Journal of Agricultural, Biological, and Environmental Statistics 14*(4), 375–391.

Madsen, L. and Y. Fang (2011). Joint regression analysis for discrete longitudinal data. *Biometrics 67*(3), 1171–1175.

Nelsen, R. B. (2006). *An Introduction to Copulas (Second edition)*. Springer New York, NY.

Nikoloulopoulos, A. K. (2013). On the estimation of normal copula discrete regression models using the continuous extension and simulated likelihood. *Journal of Statistical Planning and Inference 143*(11), 1923–1937.

Nikoloulopoulos, A. K. (2016). Efficient estimation of high-dimensional multivariate normal copula models with discrete spatial responses. *Stochastic Environmental Research and Risk Assessment 30*(2), 493–505.

Rémillard, B., N. Papageorgiou, and F. Soustra (2012). Copula-based semiparametric models for multivariate time series. *Journal of Multivariate Analysis 110*, 30–42.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.

Song, P. X.-K. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics 27*(2), 305–320.

Song, P. X.-K., M. Li, and Y. Yuan (2009). Joint regression analysis of correlated data using gaussian copulas. *Biometrics 65*(1), 60–68.

Tang, Y., H. J. Wang, Y. Sun, and A. S. Hering (2019). Copula-based semiparametric models for spatiotemporal data. *Biometrics 75*(4), 1156–1167.

Trivedi, P. and D. Zimmer (2017). A note on identification of bivariate copulas for discrete count data. *Econometrics 5*(1), 10.

Weiß, C. H., F. Zhu, and A. Hoshiyar (2022). Softplus ingarch models. *Statistica Sinica 32*(3).

Yang, L., E. W. Frees, and Z. Zhang (2020). Nonparametric estimation of copula regression models with discrete outcomes. *Journal of the American Statistical Association 115*(530), 707–720.

Department of Statistics, George Washington University

E-mail: mingze17@gwu.edu

Department of Statistics, George Washington University

E-mail: judywang@gwu.edu

Center for Statistical Research and Methodology, United States Census Bureau

E-mail: james.a.livsey@census.gov