Summer 2019

Copula-Based Zero-Inflated Count Time Series Models

Mohammed Sulaiman Alqawba
Old Dominion University, m.alqawba@qu.edu.sa

COPULA-BASED ZERO-INFLATED COUNT TIME SERIES MODELS

by

Mohammed Sulaiman Alqawba
B.S. July 2011, Qassim University, Saudi Arabia
M.S. May 2016, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

MATHEMATICS AND STATISTICS

OLD DOMINION UNIVERSITY
August 2019

Approved by:

Norou Diawara (Director)

N. Rao Chaganty (Member)

Lucia Tabacu (Member)

Khan M. Iftekharuddin (Member)

ABSTRACT

COPULA-BASED ZERO-INFLATED
COUNT TIME SERIES MODELS

Mohammed Sulaiman Alqawba
Old Dominion University, 2019
Director: Dr. Norou Diawara

Count time series data are observed in several applied disciplines such as in environmental science, biostatistics, economics, public health, and finance. In some cases, a specific count, say zero, may occur more often than usual. Additionally, serial dependence might be found among these counts if they are recorded over time. Overlooking the frequent occurrence of zeros and the serial dependence could lead to false inference. In this dissertation, we propose two classes of copula-based time series models for zero-inflated counts with the presence of covariates. Zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), and zero-inflated Conway-Maxwell-Poisson (ZICMP) distributed marginals of the counts will be considered.

For the first class, the joint distribution is modeled under Gaussian copula with autoregression moving average (ARMA) errors. Relationship between the autocorrelation function of the zero-inflated counts and the errors is studied. Sequential sampling likelihood inference is performed. To evaluate the proposed method, simulated and real-life data examples are provided and studied. For the second class, Markov zero-inflated count time series models based on a joint distribution on consecutive observations are proposed. The joint distribution function of the consecutive observations is constructed through copula functions. First or second order Markov chains are considered with the univariate margins of ZIP, ZINB, or ZICMP distributions. Under the Markov models, bivariate copula functions such as the bivariate Gaussian, Frank, and Gumbel are chosen to construct a bivariate distribution of two consecutive observations. Moreover, the trivariate Gaussian and max-infinitely divisible copula functions are considered to build the joint distribution of three consecutive observations. Likelihood based inference is performed, score functions are derived, and asymptotic properties are studied. Model diagnostic and prediction are presented. To evaluate the proposed method, simulated and real-life data examples are studied.

This dissertation is dedicated to my parents.

ACKNOWLEDGEMENTS

Reaching this point of my education would not be possible without the support, guidance, and inspiration of so many people. Without them, this dissertation and much more would not possible, and hence I would like to take this opportunity to express my gratitude.

First, I would like to thank my advisor Dr. Norou Diawara. His guidance, mentorship, and pertinence throughout my Ph.D. have been phenomenal. This dissertation and further work would not be possible without him. He opened the world of statistical application to me and introduced me to many collaborators to develop and enrich my statistical consultation skills. I will forever be in debt to him.

I am also grateful to my committee member and the statistical program director Dr. N. Rao Chaganty. His excellent feedbacks have improve this dissertation greatly, and his advising throughout my graduate studies shaped where I am today. He is an inspiration in and out of the classroom. I cannot wait to share his stories with my students. Hopefully, one day I will be as positively influential to my students as he is to me. A special thank you to my committee member Dr. Lucia Tabacu. Her invaluable comments have progressed this work. In fact, the desire to work on this subject started with her after taking the time series class with her. She was excellent lecturer and has her door always open to discuss further problems in that field. I would also like to thank my committee member Dr. Khan M. Iftekharuddin for giving us his valuable time to serve on my committee. His insightful expertise and suggestions have improved this dissertation.

Additionally, I am thankful to the Department of Mathematics and Statistics for providing such efficient work environment during my graduate studies. A special thank to the department chair Dr. Hideaki Kaneko, the graduate program director Dr. Raymond Cheng, office managers Ms. Sheila Hegwood, Ms. Miriam Venable, and my graduate collogues for their kind support throughout the last five years.

I am also thankful to the Qassim University in Saudi Arabia for providing me full financial support through my graduate studies. A special thank to my mentor Dr. Suwailam Ghanem, the Department of Mathematics' Chair at Alrass College of Sciences and Arts for giving me this chance to complete my education. I cannot wait to come back and work alongside him.

Most importantly I would like to try and express my gratitude to my family. My mother Aisha Aldossary, My father Sulaiman Alqawba, my wife Hissah Alkuraydaa, my sisters Ashwaq, Ahlam, Alaa, and my brothers Abdurahman and Ammar. There are no words to describe how important their roles on my success. Their support, guidance, and inspiration have been marvelous.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND MOTIVATION

Count time series data are observed in several applied disciplines such as in environmental science, biostatistics, economics, public health, and finance. In some cases, a specific count, say zero, may occur more often than usual. For example, for monthly counts of sandstorms in some areas, rare diseases with low infection rates, and crimes such as arson, the observed counts may include a considerable frequency of zeros. However, during certain periods, these counts could take larger values. Additionally, in practice these zero-inflated counts usually observe serial dependence when the data is collected over time. Overlooking the frequent occurrence of zeros and the serial dependence could lead to false inference. In many real-life time series examples, the series are not stationary and observe some sort of trend and seasonal features. Figure 1 shows an example of such time series. It shows the monthly counts of strong sandstorms recorded by the AQI airport station in Eastern Province, Saudi Arabia. One can see: (a) the frequent occurrence of zeros in the distribution of the counts, (b) serial dependence, (c) decreasing trend, and (d) seasonality in the series. Standard time series models fail to account for such problems. Motivated by these problems, we propose and develop two classes of copula-based time series models for zero-inflated counts with the presence of covariates. Copula is a multivariate distribution with uniform margins and allows modeling the dependence structure separately from the univariate marginal distributions.

There is a vast literature on modeling zero-inflated counts. Lambert (1992) was the first to model these types of counts via generalized linear model (GLM) assuming the counts follow zero-inflated Poisson (ZIP) distribution. Later on, other distributional assumptions such as the zero-inflated negative binomial (ZINB) distribution in Ridout et al. (2001) and zero-inflated Conway-Maxwell-Poisson (ZICMP) distribution in Sellers and Raim (2016) were proposed. With dependence structure, much of the research of dependent zero-inflated counts is on longitudinal and clustered data analysis (see for examples, Hall and Zhang (2004), Buu et al. (2012), and Choo-Wosoba and Datta (2018)).

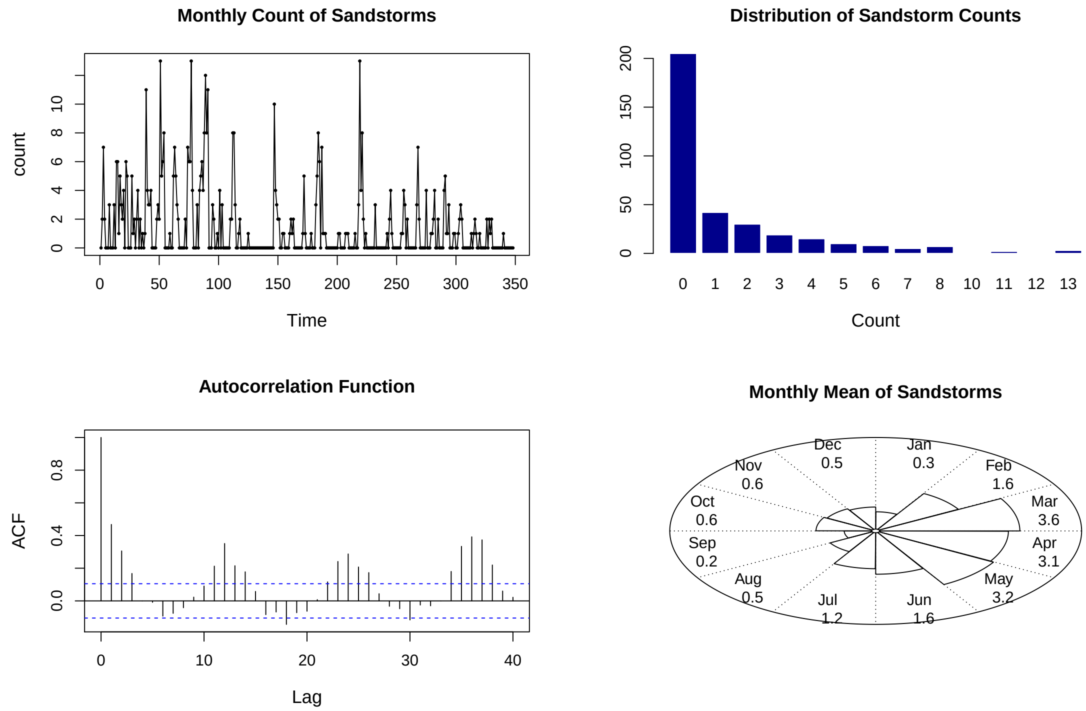


Figure 1: Time series plot of monthly count of sandstorms, bar-plot of distribution of sandstorm counts, autocorrelation function, and the circular plot.

However, recently growing interest on modeling time series of zero-inflated counts has emerged. To analyze air pollution related emergency room visit counts with excessive zeros, Hasan et al. (2012) proposed a three level random effect ZIP model. Under another class of time series models, the integer-valued generalized autoregressive conditional heteroscedatic (INGARCH), Zhu (2012) introduced ZIP and ZINB INGARCH models to fit zero-inflated time series of counts. Gonçalves et al. (2016) also introduced zero-inflated INGARCH model but with compound Poisson distribution. In state-space representation, Yang et al. (2015) proposed dynamic models to handle time series of zero-inflated counts with both ZIP and ZINB distributions. Yau et al. (2004) modeled zero-inflated time series counts of work place injuries using a mixed autoregressive model. They assumed the data followed the ZIP distribution and a correlation structure that is of first ordered autoregressive process.

There is a vast literature on modeling zero-inflated counts. Lambert (1992) was the first to model these types of counts via generalized linear model (GLM) assuming the counts follow zero-inflated Poisson (ZIP) distribution. Later on, other distributional assumptions such as the zero-inflated negative binomial (ZINB) distribution in Ridout et al. (2001)

and zero-inflated Conway-Maxwell-Poisson (ZICMP) distribution in Sellers and Raim (2016) were proposed. With dependence structure, much of the research of dependent zero-inflated counts is on longitudinal and clustered data analysis (see for examples, Hall and Zhang (2004), Buu et al. (2012), and Choo-Wosoba and Datta (2018)). However, recently growing interest on modeling time series of zero-inflated counts has emerged. To analyze air pollution related emergency room visit counts with excessive zeros, Hasan et al. (2012) proposed a three level random effect ZIP model. Under another class of time series models, the integer-valued generalized autoregressive conditional heteroscedatic (INGARCH), Zhu (2012) introduced ZIP and ZINB INGARCH models to fit zero-inflated time series of counts. Gonçalves et al. (2016) also introduced zero-inflated INGARCH model but with compound Poisson distribution. In state-space representation, Yang et al. (2015) proposed dynamic models to handle time series of zero-inflated counts with both ZIP and ZINB distributions. Yau et al. (2004) modeled zero-inflated time series counts of work place injuries using a mixed autoregressive model. They assumed the data followed the ZIP distribution and a correlation structure that is of first ordered autoregressive process.

To add flexibility in the correlation structure of continuous time series data, copula has been proposed by many authors (see for examples, Joe 2014, Guolo and Varin 2014, and Patton 2009). However, there is not much literature for count time series data as there is for continuous measurements due to computational complexity. Joe (2016) suggested copula-based Markov model for time series of counts. Masarotto and Varin (2012) introduced marginal regression models for count time series data where the serial dependence being captured by a Gaussian copula, with a correlation matrix corresponding to a stationary autoregressive moving average (ARMA) process. They performed statistical inference through an approximated likelihood function using sequential importance sampling technique. Lennon (2016) and Jia et al. (2018) also applied the same models but suggested different estimation methods including approximate Bayesian computation of the likelihood function and pseudo Gaussian likelihood estimation, respectively.

1.1 OVERVIEW OF THE DISSERTATION

In this dissertation, we propose and develop two classes of copula-based zero-inflated count time series models. In Chapter 2, we review the background and the literature needed to construct the proposed two classes of models. In particular, we provide an overview of the time series literature in both the continuous and discrete cases and discuss why some of these models fail in handling count time series with excess zeros. Next, we review some of the zero-inflated count regression models applied to the marginal distributions. The chosen marginal distributions are the ZIP, ZINB, and ZICMP margins. Then, we present the copula theory needed to build the dependence structures of our models.

In Chapter 3, we extend the work done in Masarotto and Varin (2012) by including a class with zero-inflated distributions such as the ZIP, ZINB and ZICMP distributions whereas the joint distribution is modeled under Gaussian copula with autoregression moving average (ARMA) errors. Relationship between the autocorrelation function of the zero-inflated counts and the errors is studied. Sequential importance sampling likelihood inference is performed to estimate both the marginal regression parameters and the dependence (or copula) parameters. We run residual analysis to evaluate the performance of the models. To evaluate the proposed method in this chapter, simulation studies are conducted. We apply the proposed models on two real-life examples from two different areas to show the potential of such class of models.

In Chapter 4, we introduce a class of Markov models similar to the one in Joe (2016) but with zero-inflated margins. The chapter concentrates on building a class of Markov zero-inflated count time series models based on a joint distribution of consecutive observations. The joint distribution function of the consecutive observations is constructed through copula functions such as the Gaussian, Frank, and Gumbel copula functions. We list some of the model properties and review how to measure the dependence structure when the chosen copula function is not Gaussian. We perform maximum likelihood estimation method, derive score functions, and show asymptotic properties using results from Billingsley (1961). Model selection and prediction are implemented to assess the performance of the models. Simulation studies are also conducted to evaluate the estimation method and the asymptotic behavior of the parameter estimates. Finally, we apply the proposed models on two real-life examples from two different areas.

In Chapter 5, we summarize the work the dissertation and discuss future directions.

Some results and selected R codes are provided in Appendices A and B.

CHAPTER 2

BACKGROUND

2.1 TIME SERIES ANALYSIS

A time series is a sequence of random variables, $\{Y_t\}$, taken over an equally spaced discrete time, that is, $t = 1, \ldots, n$ for some fixed n. The observed values, $\{y_t\}$ are usually referred as the realization of the stochastic process $\{Y_t\}$. An important feature of a time series is that consecutive observations are serially dependent. Hence, time series analysis concerns with accounting for the serial dependence. A complete description of a time series process is the joint multivariate distribution of the observed values, that is

$$F(y_1, \ldots, y_n) = \text{Pr}(Y_1 \leq y_1, \ldots, Y_n \leq y_n). \tag{1}$$

However, constructing these multivariate distributions is quite challenging especially when the time series is discrete and observed an unusual behavior such as the zero inflation. As mentioned in the introduction, the main objective of this dissertation is to build a multivariate distribution as given in (1) to describe the zero-inflated count time series through two different classes of copula based models.

Important definitions in the time series context are given without making any assumptions. First, we will define both strictly and weakly stationary time series as given in Shumway and Stoffer (2011).

Definition 1. *A strictly stationary time series is one for which the probabilistic behavior of every collection of values*

$$\{Y_1, \ldots, Y_k\}$$

is identical to that of the time shifted set

$$\{Y_{1+h}, \ldots, Y_{k+h}\}.$$

That is,

$$Pr(Y_1 \leq y_1, \ldots, Y_k \leq y_k) = Pr(Y_{1+h} \leq y_{1+h}, \ldots, Y_{k+h} \leq y_{k+h})$$

for all $k = 1, \ldots, n$*, and all* $h = 0, \pm 1, \pm 2, \ldots$*.*

The strictly stationary condition in Definition 1 is quite strong for most applications. Hence, weaker condition has been imposed on time series data. The definition of the weak condition is the following.

Definition 2. *A weakly stationary time series,* Y_t*, is a finite variance process such that*

1. *the mean* $\mu = E(Y_t)$ *is constant and does not depend on time* t*, and*
2. *the autocovariance function*

   $$\gamma(h) = Cov(Y_t, Y_{t+h})$$

   depends on the time, t*, through only* h*.*

Next, we review some statistical models for analyzing time series and stochastic processes. We start with one of the most popular processes, the autoregressive moving average (ARMA) process in Subsection 2.1.1. We define the process and discuss its properties, and why the analysis for such processes fails when the observations of the time series are discrete. In Subsection 2.1.2, we review several statistical models that are appropriate for handling count time series. We start with the discrete version of the ARMA process, and then discuss how modeling count time series evolved over time with thinning-operators based models and GLM-based models.

2.1.1 AUTOREGRESSIVE MOVING AVERAGE (ARMA) MODELS

One of the most popular linear processes to fit a stationary time series process is the ARMA process. This type of process is constructed through a combination of two process, the autoregressive (AR) process and the moving average (MA) process. That is, assume $\{\epsilon_t\}$ is a stationary time series process. Then, an autoregressive process of order p ($p \geq 0$) would be modeled as

$$\epsilon_t = \sum_{i=1}^{p} \varphi_i \epsilon_{t-i} + \eta_t; \qquad \eta_t \sim WN(0, \sigma_\eta^2)$$

where WN represents a stationary white noise, and $\varphi_1, \ldots, \varphi_p$ are constant ($\varphi_p \neq 0$). On the other hand, the moving average process of order q ($q \geq 0$) would be modeled as

$$\epsilon_t = \sum_{j=1}^{q} \delta_j \eta_{t-j} + \eta_t; \qquad \eta_t \sim WN(0, \sigma_\eta^2)$$

where $\delta_1, \ldots, \delta_q$ are constants ($\delta_q \neq 0$). Hence, combining the above two processes yields to the autoregressive moving average processes of order (p, q), i.e. $ARMA(p, q)$ and it is given by.

$$\epsilon_t = \sum_{i=1}^{p} \varphi_i \epsilon_{t-i} + \sum_{j=1}^{q} \delta_j \eta_{t-j} + \eta_t. \tag{2}$$

In most applications, the white noise in (2) follows the Gaussian distribution, i.e. $\eta_t \sim N(0, \sigma_\eta^2)$, so the process would be called Gaussian ARMA, which will play an important role in the copula based model presented in Chapter 3.

The ARMA process can also be represented as

$$\varphi(B)\epsilon_t = \delta(B)\eta_t,$$

where $\varphi(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i, \delta(B) = 1 + \sum_{j=1}^{q} \delta B^j$ and B is the back shift operator such that $B\epsilon_t = \epsilon_{t-1}$ and $B^i \epsilon_t = \epsilon_{t-i}$. Hence, important characteristics of the ARMA process can be derived, which are the causality and invertibility.

An $ARMA(p, q)$ process is said to be causal, if it can be expressed as a moving average process, i.e.

$$\psi(B)\epsilon_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} = \eta_t,$$

where $\psi(B) = \sum_{i=1}^{\infty} \psi_i B^i$, and $\{\psi_i\}$ is a given sequence of constants given by solving $\psi(B) = \delta(B)/\varphi(B)$, and satisfying $\sum_{i=1}^{\infty} |\psi_i| < \infty$; we set $\psi_0 = 1$.

An $ARMA(p, q)$ process is said to be invertible, if it can be expressed as

$$\epsilon_t = \sum_{i=0}^{\infty} \pi_i \eta_{t-i} = \pi(B)\eta_t,$$

where $\pi(B) = \sum_{i=1}^{\infty} \pi_i B^i$, and $\{\pi_i\}$ is a given sequence of constants given by solving

$\psi(B) = \varphi(B)/\delta(B)$, and satisfying $\sum_{i=1}^{\infty}|\pi_i| < \infty$; we set $\pi_0 = 1$.

Obtaining the autocovariance, $\gamma(h)$ of an $ARMA(p, q)$ process required recursive solutions of homogeneous equations, known as the difference equations. The general homogeneous equation for the autocovariance of a causal ARMA process is given by

$$\gamma(h) - \varphi_1\gamma(h-1) - \cdots - \varphi_p\gamma(h-p) = 0, \tag{3}$$

for $h \geq \max(p, q+1)$ with initial conditions

$$\gamma(h) - \sum_{i=1}^{p}\varphi_i\gamma(h-i) = \sigma_\eta^2\sum_{j=h}^{q}\delta_j\psi_{j-h},$$

for $0 \leq h \leq \max(p, q+1)$. When $h = 0$, we have

$$\gamma(0) - \sum_{i=1}^{p}\varphi_i\gamma(i) = \sigma_\eta^2\sum_{j=h}^{q}\delta_j\psi_j,$$

so that

$$\sigma_\eta^2 = \frac{\gamma(0) - \sum_{i=1}^{p}\varphi_i\gamma(i)}{\sum_{j=h}^{q}\delta_j\psi_j}, \tag{4}$$

where the coefficients ψ_h can be found recursively by

$$\psi_h = \varphi_h + \delta_h + \delta_{h-1}\psi_1 + \delta_{h-2}\psi_2 + \cdots + \delta_1\psi_{h-1}$$

where $\varphi_h = 0$ for $h > p$, $\delta_h = 0$ for $h > q$, and $\psi_0 = 1$. Dividing (3) through by $\gamma(0)$ gives a similar recursion on $\rho(h) = \gamma(h)/\gamma(0)$, the autocorrelation function. More details on the ARMA process and its properties can be found on Brockwell and Davis (2013) and Box et al. (1994) for examples.

2.1.2 DISCRETE TIME SERIES MODELS

Analyzing and modeling discrete time series data have drawn interests because of the challenges related to the discreteness of the data. Such data violate the assumptions made on typical linear models such as ARMA models. In this subsection, we review several statistical classes models that are appropriate for handling discrete time series.

Comprehensive reviews of discrete time series can be found in MacDonald and Zucchini (1997), McKenzie (2003), Kedem and Fokianos (2005), and Davis et al. (2016).

The first attempt to provide a class of models that fit discrete time series data was introduced by Jacobs and Lewis (1978a,b,c). Their models are similar to the ARMA process and are called discrete ARMA (DARMA). They provide a stationary time series of dependent random variables with marginal distributions and dependence structure specified independently. Although the DARMA class of models is very attractive in theory, its construction is unusual for applications (McKenzie, 2003). In addition, the models are only capable of providing positive autocorrelation. Most of the applications of these models are on the hydrological literature (see for examples, Buishand, 1978, Chang et al., 1984b,a, 1987, and Delleur et al., 1989).

Another popular class of models that fits count time series data is the one in which models are based on the idea of thinning operators. Such models and their properties are studied extensively by many authors. McKenzie (2003), Weiß (2008), and Joe (2016) reviewed a variety of classes of count time series data based on thinning operators. Different marginal distributions and thinning operators were discussed in Steutel and Van Harn (1979), McKenzie (1985, 1986, 1987, 1988), Al-Osh and Alzaid (1987, 1988, 1991), Alzaid and Al-Osh (1988, 1990, 1993), Zhu and Joe (2006), and Zhu and Joe (2010a,b). Generalized linear models (GLMs) have been also introduced to model time series data with added covariates using regression setting. Fokianos (2015) reviewed such models under different distributional assumptions and links functions. Although the Poisson distribution is traditionally the first choice to model count data, other researchers have been considering alternative distributions. The most common alternative to model counts is the negative binomial (NB) distribution, which also allows for overdispersion in the data. Davis and Wu (2009) proposed a similar model to the Poisson model with logarithmic link function but with NB distribution and logit link function. Chen et al. (2016) also considered the NB distribution for modeling count time series data but with an autoregressive conditional model. Zhu (2011) proposed a negative binomial $INGARCH(p, q)$ process denoted as $NBINGARCH(p, q)$. Yang et al. (2013) applied similar approach with zero-inflated distributions, specifically, the ZIP and ZINB distribution.

Most of the above models fails to incorporate certain marginal distributions without facing complications and violating important assumptions. Furthermore, including covariates and accounting for non-stationarity are challenging and sometimes not possible with the

framework of some of the models mentioned in this section. Next, we review zero-inflated regression models and copula theory. We conclude this chapter by brief overview of copula-based count time series models and how they can be extended to zero-inflated counts, which is the contribution of this dissertation.

2.2 ZERO-INFLATED COUNT REGRESSION MODELS

In this section, we will revisit the zero inflated regression models under different distributional assumptions. First, we will consider the zero-inflated Poisson regression model with independent counts (Lambert, 1992). Then other distributional assumptions such as the zero-inflated negative binomial distribution (Ridout et al., 2001), and the zero-inflated Conway-Maxwell-Poisson (Sellers and Raim, 2016) are considered.

2.2.1 ZERO-INFLATED POISSON DISTRIBUTION

Suppose Y_t denotes a random count at time t. We say that Y_t is distributed as $ZIP(\omega_t, \lambda_t)$, the zero-inflated Poisson distribution with parameters ω_t and λ_t, if

$$Y_t \sim \begin{cases} 0, & \text{with probability } \ \omega_t, \\ \text{Poisson}(\lambda_t), & \text{with probability } \ 1-\omega_t, \end{cases} \tag{5}$$

for $t = 1, \dots, n$, so that

$$Y_t = \begin{cases} 0, & \text{with probability } \ \omega_t + (1-\omega_t)e^{-\lambda_t}, \\ y_t, & \text{with probability } \ (1-\omega_t)e^{-\lambda_t}\lambda_t^{y_t}/y_t!, \ \ y_t = 1, 2, \dots, \end{cases}$$

where $\omega_t \in [0, 1]$ is the zero-inflation parameter as often referred to, and $\lambda_t > 0$ is the intensity parameter, or the mean, of the baseline Poisson distribution . The probability mass function (pmf) of the ZIP distribution can also be written as

$$f_{Y_t}(y_t) = \omega_t I_{\{y_t=0\}} + (1-\omega_t)\frac{e^{-\lambda_t}\lambda_t^{y_t}}{y_t!}, \tag{6}$$

where $I_{\{y_t=0\}}$ is in an indicator function.

Note that (5) is a mixture of a degenerate distribution with point mass at zero and Poisson distribution. Hence, defining a new binary variable, say v_t, that takes the value zero if y_t comes from the Poisson distribution and the value one if y_t comes from the degenerate distribution, we have

$$v_t \sim \text{Bernoulli}(\omega_t), \tag{7}$$

so that

$$Y_t|v_t \sim \text{Poisson}((1-v_t)\lambda_t), \tag{8}$$

where $v_t's$ are assumed to be independent. From (7) and (8), we can see that if ω_t is zero, the ZIP distribution becomes the ordinary Poisson distribution.

Furthermore, for any integer $y_t \geq 0$, the cumulative distribution function (cdf) of Y_t for $t = 1, \ldots, n$ is given by

$$\begin{aligned} F_{Y_t}(y_t) &= \text{Pr}(Y_t \leq y_t) \\ &= \sum_{m=0}^{y_t} \text{Pr}(Y_t = m) \\ &= \omega_t + (1-\omega_t)e^{-\lambda_t} \sum_{m=0}^{y_t} \frac{\lambda_t^m}{m!}. \end{aligned} \tag{9}$$

The first two moments can be derived using (7) and (8). The mean of Y_t is given by

$$\text{E}(Y_t) = \text{E}[\text{E}(Y_t|v_t)] = \text{E}[(1-v_t)\lambda_t] = (1-\omega_t)\lambda_t,$$

and the variance of Y_t is given by:

$$\begin{aligned} \text{Var}(Y_t) &= \text{E}[\text{Var}(Y_t|v_t)] + \text{Var}[\text{E}(Y_t|v_t)] \\ &= \text{E}[(1-v_t)\lambda_t] + \text{Var}[(1-v_t)\lambda_t] \\ &= (1-\omega_t)(1+\omega_t\lambda_t)\lambda_t. \end{aligned} \tag{10}$$

Notice that unless $\omega_t = 0$, the variance is greater than the mean. Thus, the zero inflation results in overdispersion, which cannot be captured by the ordinary Poisson distribution.

Following GLM (Nelder and Wedderburn, 1972), the regression model to the ZIP distributed random variables is fitted using the proposed method by Lambert (1992). The idea is to simultaneously fit two GLM's for: 1) the intensity parameter λ_t, with the logarithmic link function, and 2) the zero-inflation parameter ω_t for $t = 1, \ldots, n$ with the logit link function, which are given by

$$\log(\lambda_t) = \boldsymbol{\beta}'\mathbf{x}_t, \tag{11}$$

and

$$\text{logit}(\omega_t) = \log\left(\frac{\omega_t}{1-\omega_t}\right) = \boldsymbol{\gamma}'\mathbf{z}_t, \tag{12}$$

where $\mathbf{x}_t = (x_{1t}, \ldots, x_{kt})'$ and $\mathbf{z}_t = (z_{1t}, \ldots, z_{lt})'$ are the associated covariates that affect the intensity parameter λ_t and the zero-inflation parameter ω_t, respectively. In addition, $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_k)$ and $\boldsymbol{\gamma}' = (\gamma_1, \ldots, \gamma_l)$ are the regression coefficients for the log-linear model given in (11) and the logit model given in (12), respectively. Note that, $\mathbf{x}_t$ and $\mathbf{z}_t$ are not necessarily the same. Clearly when they are the same, this ZIP regression model has twice as many parameters as the ordinary Poisson regression model. However, in the case where the zero-inflation parameter does not depend on covariates, then vector reduce to a scaler that takes the value one, i.e. $z_t = 1$ for $t = 1, \ldots, n$ (Lambert, 1992).

2.2.2 ZERO-INFLATED NEGATIVE BINOMIAL DISTRIBUTION

Although the ZIP distribution is the most popular candidate to handle zero-inflations, other distributions are also considered in the literature due to the inflexibility associated with the ZIP distribution. One of the drawbacks of ZIP is that it cannot deal with over-dispersion especially among the non-zero values that come from the Poisson distribution. And when the counts are dependent, the ZIP parameter estimates can be biased (Yau et al., 2003). An alternative is to replace the one parameter Poisson distribution with the two parameter negative binomial distribution resulting in ZINB distribution.

The ZINB distribution (or the NB distribution) has always been the natural second choice after the ZIP distribution (or the Poisson distribution) since it has the ability to handle the over-dispersion among the count data. We say Y_t is distributed as $ZINB(\omega_t, \lambda_t, \kappa_t)$, the zero inflated negative binomial with parameters ω_t, λ_t and κ_t if the pmf is given by

$$f_{Y_t}(y_t) = \omega_t I_{\{y_t=0\}} + (1-\omega_t)\frac{\Gamma(\kappa_t + y_t)}{\Gamma(\kappa_t) y_t!}\left(\frac{\kappa_t}{\kappa_t+\lambda_t}\right)^{\kappa_t}\left(\frac{\lambda_t}{\kappa_t+\lambda_t}\right)^{y_t}, \tag{13}$$

where $\omega_t \in [0,1]$ is the zero-inflation parameter, $\lambda_t > 0$ is the intensity parameter, or the mean, of the baseline negative binomial distribution, and finally $\kappa_t \geq 0$ is the dispersion parameter.

For any integer $y_t \geq 0$, the cdf of Y_t for $t = 1, \ldots, n$ is given by

$$F_{Y_t}(y_t) = \omega_t + \frac{(1-\omega_t)}{\Gamma(\kappa_t)}\Big(\frac{\kappa_t}{\kappa_t+\lambda_t}\Big)^{\kappa_t}\sum_{m=0}^{y_t}\frac{\Gamma(\kappa_t+m)}{m!}\Big(\frac{\lambda_t}{\kappa_t+\lambda_t}\Big)^{m}. \tag{14}$$

The moments can be derived via a latent variable as with the ZIP distribution. Let v_t for $t = 1, \ldots, n$ takes the value zero if y_t comes from the NB distribution and the value one if y_t comes from the degenerate distribution at point mass zero. Then, the latent variable v_t is distributed as

$$v_t \sim \text{Bernoulli}(\omega_t),$$

so that

$$Y_t|v_t \sim \text{NB}(\kappa_t, (1-v_t)\lambda_t).$$

Then based on this hierarchical representation, the mean of the ZINB distributed random variable Y_t is given by:

$$\text{E}(Y_t) = \text{E}[\text{E}(Y_t|v_t)] = (1-\omega_t)\lambda_t,$$

and the variance of Y_t is given by:

$$\begin{aligned}\text{Var}(Y_t) &= \text{E}[\text{Var}(Y_t|v_t)] + \text{Var}[\text{E}(Y_t|v_t)] \\ &= (1-\omega_t)(1+\omega_t\lambda_t+\lambda_t/\kappa_t)\lambda_t. \end{aligned} \tag{15}$$

Here, regardless of ω_t value, the variance of Y_t is always greater than the mean. This suggests that the overdispersion if found in the count data would be always captured by the ZINB distribution.

We will simultaneously fit three GLM's for: 1) the intensity parameter λ_t, with the logarithmic link function, 2) the zero-inflation parameter ω_t for $t = 1, \ldots, n$ with the logit link function, and finally 3) the dispersion parameter κ_t, with the logarithmic link function

just like the intensity parameter. That is,

$$\log(\lambda_t) = \boldsymbol{\beta}'\mathbf{x}_t, \tag{16}$$

$$\text{logit}(\omega_t) = \log\left(\frac{\omega_t}{1-\omega_t}\right) = \boldsymbol{\gamma}'\mathbf{z}_t, \tag{17}$$

and

$$\log(\kappa_t) = \boldsymbol{\alpha}'\mathbf{w}_t, \tag{18}$$

where $\mathbf{x}_t = (x_{1t}, \ldots, x_{kt})'$, $\mathbf{z}_t = (z_{1t}, \ldots, z_{lt})'$ and $\mathbf{w}_t = (w_{1t}, \ldots, w_{mt})'$ are the associated covariates that affect the intensity parameter λ_t, the zero-inflation parameter ω_t, and the dispersion parameter κ_t, respectively. In addition, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_l)'$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)'$ are the regression coefficients for the log-linear model given in (16), the logit model given in (17) and the log-linear model given in (18), respectively.

2.2.3 ZERO-INFLATED CONWAY-MAXWELL-POISSON DISTRIBUTION

Recently, Sellers and Raim (2016) have introduced ZICMP model, which essentially replaces the Poisson distribution in the ZIP model by the Conway-Maxwell-Poisson (CMP) distribution (Conway and Maxwell, 1962; Shmueli et al., 2005). Note that ZICMP is a generalization of ZIP since the Poisson distribution is a special case of CMP distribution. An added advantage of ZICMP is that it not only can handle over-dispersion but also under-dispersion in the counts. The ZICMP is highly flexible because it allows the dispersion to be in any direction.

The probability mass function of the zero inflated CMP distribution with parameters ω_t , λ_t and κ_t ($ZICMP(\omega_t, \lambda_t, \kappa_t)$) is given by

$$f_{Y_t}(y_t) = \omega_t I_{\{y_t=0\}} + (1-\omega_t)\frac{\lambda_t^{y_t}}{(y_t!\,)^{\kappa_t} Z(\lambda_t, \kappa_t)}, \tag{19}$$

where as before $\omega_t \in [0, 1]$ is the zero-inflation parameter, and $\lambda_t \geq 0, \kappa_t \geq 0$ are the parameters of the underlying CMP distribution. The infinite series $Z(\lambda_t, \kappa_t) = \sum_{j=0}^{\infty} \frac{\lambda_t^j}{(j!)^{\kappa_t}}$ is the normalizing function of the CMP distribution.

For any integer $y_t \geq 0$, the cdf of Y_t for $t = 1, \ldots, n$ is given by

$$F_{Y_t}(y_t) = \omega_t + \frac{(1-\omega_t)}{Z(\lambda_t, \kappa_t)} \sum_{m=0}^{y_t} \frac{\lambda_t^m}{(m!\,)^{\kappa_t}}. \tag{20}$$

The moments of the CMP distribution are not in a closed formed. However approximations have been suggested (see Shmueli et al. 2005 for more details). Assuming that v_t is distributed as $CMP(\lambda_t, \kappa_t)$ for $t = 1, \ldots, n$ with the first and second moments given in Shmueli et al. (2005) by

$$\mathrm{E}(v_t) = \lambda_t \frac{\partial \log Z(\lambda_t, \kappa_t)}{\partial \lambda_t},$$

$$\mathrm{E}(v_t^2) = \frac{\lambda_t \partial \mathrm{E}(v_t)}{\partial \lambda_t} + [\mathrm{E}(v_t)]^2.$$

Hence, the expected value of the ZICMP is given by

$$\mathrm{E}(Y_t) = (1-\omega_t)\mathrm{E}(v_t),$$

and the variance

$$\begin{aligned}
\mathrm{Var}(Y_t) &= \mathrm{E}(Y_t^2) - [\mathrm{E}(Y_t)]^2 \\
&= (1-\omega_t)\mathrm{E}(v_t^2) - (1-\omega_t)^2[\mathrm{E}(v_t)]^2 \\
&= (1-\omega_t)\frac{\lambda_t \partial \mathrm{E}(v_t)}{\partial \lambda_t} + (1-\omega_t)[\mathrm{E}(v_t)]^2 - (1-\omega_t)^2[\mathrm{E}(v_t)]^2 \\
&= (1-\omega_t)\Big(\frac{\lambda_t \partial \mathrm{E}(v_t)}{\partial \lambda_t} + \omega_t[\mathrm{E}(v_t)]^2\Big),
\end{aligned}$$

where the expected of the CMP, i.e. $\mathrm{E}(v_t)$ is approximated by

$$\mathrm{E}(v_t) \approx \lambda_t^{1/\kappa_t} - \frac{\kappa_t - 1}{2\kappa_t}.$$

Using this approximation, one can approximate the expected value and variance of the ZICMP by

$$\mathrm{E}(Y_t) \approx (1-\omega_t)\Big(\lambda_t^{1/\kappa_t} - \frac{\kappa_t - 1}{2\kappa_t}\Big),$$

$$\mathrm{Var}(Y_t) \approx (1-\omega_t)\Big[\frac{\lambda_t^{1/\kappa_t}}{\kappa_t} + \omega_t\Big(\lambda_t^{1/\kappa_t} - \frac{\kappa_t - 1}{2\kappa_t}\Big)^2\Big]. \tag{21}$$

Similar to the ZINB, the ZICMP regression model uses the same link functions given in (16), (18) and (17). In the next section, we will describe the use of copulas to construct multivariate distributions for a time series of count variables with given zero-inflated marginal distributions.

2.3 COPULAS

The multivariate normal distribution is a natural extension of the univariate normal distribution to higher dimensions and it plays a central role in statistical theory. However, there is not one but several multivariate extensions of univariate discrete distributions to higher dimensions. And copulas facilitate such extensions and construction of mutivariate distributions with given continuous or discrete marginals, to model various types of dependence. An extensive and detailed discussion of copulas is contained in Joe (2014). A n dimensional copula $C(a_1, a_2, \ldots, a_n) : [0,1]^n \to [0,1]$ is simply a multivariate cdf with all n univariate marginals uniform on the unit interval, and it satisfies the following three properties.

1. $C(1, \ldots, u_t, \ldots, 1) = u_t$, for $u_t \in [0,1]$, and $t = 1, 2, \ldots, n$.
2. $C(u_1, u_2, \ldots, u_n) = 0$ if at least one $u_t = 0$ for $t = 1, 2, \ldots, n$.
3. For any $u_{t_1}, u_{t_2} \in [0,1]$ with $u_{t_1} \leq u_{t_2}$, for $t = 1, 2, \ldots, n$,

$$\sum_{j_1=1}^{2}\sum_{j_2=1}^{2}\cdots\sum_{j_n=1}^{2}(-1)^{j_1+j_2+\cdots+j_n}C(u_{1j_1}, u_{2j_2}\ldots, u_{nj_n}) \geq 0.$$

Property 3 ensures that probability of n-dimensional rectangles is non-negative. Suppose F_t is the cdf of Y_t for $t = 1, 2, \ldots, n$. Then a multivariate cdf for $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ is given by

$$F(y_1, y_2, \ldots, y_n) = C(F_1(y_1), F_2(y_2), \ldots, F_n(y_n)).$$

If the marginal distributions $F_t(\cdot)$'s are continuous with pdfs $f_t(\cdot)$'s then the density function is given by

$$f(y_1, y_2, \ldots, y_n) = c(F_1(y_1), F_2(y_2), \ldots, F_n(y_n)) \prod_{t=1}^{n} f_t(y_t), \quad y_t \in \mathbb{R},$$

where c is the copula density given by $c(\mathbf{u}) = \partial C(\mathbf{u})/\partial \mathbf{u}$, where $\mathbf{u} = (u_1, \ldots, u_n)$, and $0 \leq u_t \leq 1$ for $t = 1, \ldots, n$. When all the margins are integer valued as given on page 27, Joe (2014), the multivariate probability mass function can be obtained as

$$\begin{aligned} f(y_1, y_2, \ldots, y_n) &= P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) \\ &= \sum_{j_1=1}^{2} \sum_{j_2=1}^{2} \cdots \sum_{j_n=1}^{2} (-1)^{j_1+j_2+\cdots+j_n} C(u_{1j_1}, u_{2j_2} \ldots, u_{nj_n}), \end{aligned} \tag{22}$$

where $u_{t1} = F_t(y_t)$ and $u_{t2} = F_t(y_{t^-})$. The term $F_t(y_{t^-})$ is the left-hand limit of F_t at y_t, which is equal to $F_t(y_t - 1)$ when Y_t is integer valued random variable. A comprehensive list of copulas can be found in Nelsen (2007) and Joe (2014). Next, we review some of these that are used in this dissertation.

2.3.1 COPULA FUNCTIONS

There are numerous copulas available, and one of the most popular copulas in the literature is the Gaussian copula. The Gaussian copula shares many of the properties of multivariate normal (Gaussian) distribution such as the correlation structure. Therefore, the flexibility to manipulate the association structure by using the Gaussian copula will be taken advantage of. This copula and other copula functions are defined next.

Gaussian Copula

The copula associated with standard multivariate Gaussian distribution called Gaussian copula, and its function is given by

$$C(u_1, \ldots, u_n) = \Phi_R(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)), \quad \forall u_i \in [0, 1], \tag{23}$$

where Φ^{-1} is the inverse CDF of a standard normal and Φ_R is the joint CDF of a standard multivariate normal distribution with covariate matrix equal to the positive definite correlation matrix R. The Gaussian copula density is defined as:

$$c(u_1, \dots, u_n) = \frac{\phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n); R)}{\prod_{t=1}^{n} \phi(\Phi^{-1}(u_t))}, \quad \forall u_i \in [0, 1],$$

where ϕ is the pdf of a standard normal and ϕ_R is the joint pdf of a standard multivariate normal distribution.

When the dimension $n = 2$, the bivariate Gaussian copula with correlation parameter reduces to a scaler, ρ, is given by

$$C(u_1, u_2; \rho) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), ,$$

for $u_i \in [0, 1], i = 1, 2$ and $-1 \leq \rho \leq 1$. The density function given as

$$c(u_1, u_2; \rho) = \frac{\Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1))\phi(\Phi^{-1}(u_2))}.$$

The conditional distribution is then given by

$$C(u_1|u_2; \rho) = \Phi\Big(\frac{\Phi^{-1}(u_1) - \rho\Phi^{-1}(u_2)}{\sqrt{1 - \rho^2}}\Big).$$

Bivariate Frank Copula

The bivariate Frank copula (Frank, 1979) is given by

$$C(u_1, u_2; \delta) = \frac{1}{\delta} \log\Big(\frac{1 - e^{-\delta} - (1 - e^{-\delta u_1})(1 - e^{-\delta u_2})}{1 - e^{-\delta}}\Big),$$

for $u_i \in [0, 1], i = 1, 2$ and $-\infty < \delta < \infty$. The density is then given by

$$c(u_1, u_2; \delta) = \frac{\delta(1 - e^{-\delta})e^{-\delta(u_1 + u2)}}{1 - e^{-\delta}}.$$

The conditional distribution function of the Frank copula is given by

$$C(u_1|u_2; \delta) = e^{-\delta u_2}[(1 - e^{-\delta})(1 - e^{-\delta u_1})^{-1} - (1 - e^{-\delta u_2})]^{-1}.$$

Bivariate Plackett Copula

The bivariate Plackett copula (Plackett, 1965; Mardia, 1967) is given by

$$\begin{aligned} C(u_1, u_2; \delta) &= \frac{1}{2(\delta-1)}\{1 + (\delta-1)(u_1+u_2) - \\ &\quad [(1+(\delta-1)(u_1+u_2))^2 - 4\delta(\delta-1)u_1u_2]^{1/2}\}, \end{aligned}$$

for $u_i \in [0,1], i = 1,2$ and $0 \leq \delta < \infty$. The density is then given by

$$c(u_1, u_2; \delta) = \frac{\delta[1+(\delta-1)(u_1+u_2-2u_1u_2)]}{[(1+(\delta-1)(u_1+u_2))^2 - 4\delta(\delta-1)u_1u_2]^{3/2}}.$$

The conditional distribution function of the Plackett copula is given by

$$C(u_1|u_2; \delta) = \frac{1}{2} - \frac{1}{2}\frac{(\delta-1)u_2 + 1 - (\delta+1)u_1}{[(1+(\delta-1)(u_1+u_2))^2 - 4\delta(\delta-1)u_1u_2]^{1/2}}.$$

Bivariate Gumbel Copula

The bivariate Gumbel copula (Gumbel, 1960) is given by

$$C(u_1, u_2; \delta) = \exp\{-([-\log u_1]^\delta + [-\log u_2]^\delta)^{1/\delta}\},$$

for $u_i \in [0,1], i = 1,2$ and $1 \leq \delta < \infty$. The density is then given by

$$\begin{aligned} c(u_1, u_2; \delta) &= \exp\{-[x^\delta + y^\delta]^{1/\delta}\} \\ &\quad \times[(x^\delta + y^\delta)^{1/\delta} + \delta - 1][x^\delta + y^\delta]^{1/\delta - 2}(u_1u_2)^{-1}, \end{aligned}$$

where $x = -\log u_1, y = -\log u_2$.
The conditional distribution function of the Gumbel copula is given by

$$C(u_1|u_2; \delta) = \frac{1}{u_2}\exp\{-[x^\delta + y^\delta]^{1/\delta}\}[1 + (x/y)^\delta]^{1/\delta - 1}.$$

The bivariate *reflected or survival* Gumbel copula is given by

$$C(u_1, u_2; \delta) = u_1 + u_2 - 1 + \exp\{-([-\log u_1]^{\delta} + [-\log u_2]^{\delta})^{1/\delta}\}.$$

Mixtures of Max-id Copula

If the gaol is to construct a multivariate copula function with closed form expression, one can use the idea of mixing max-infinitely divisible (max-id) distributions (Joe and Hu, 1996). A multivariate distribution is called max-id if all its powers are positive and also proper distribution function. Let H be max-id $n-$variate copula, and let ψ be the Laplace transform (LT) function of a positive random variable. Then, a multivariate distribution is given by

$$F_{\psi,H} = \int_0^{\infty} H^{st} dF_S(s) = \psi(-\log H), \quad t \geq 0,$$

for some positive random variable $s \sim F_S$ with a copula function given by

$$\begin{aligned} &C_{\psi,H}(u_1, \ldots, u_n) \\ &= \psi\Big(- \log H(\exp\{-\psi^{-1}(u_1)\}, \ldots, \exp\{-\psi^{-1}(u_n)\})\Big), \end{aligned} \tag{24}$$

for all $u_i \in [0, 1]$. Several special cases of (24) have been studied and applied to different applications. The Archimedean copula is one example and can be given by

$$C_{\psi,H}(u_1, \ldots, u_n) = F_{\psi,H}\Big(\sum_{i=1}^{n} F_{\psi,H}^{-1}(u_i)\Big).$$

Joe and Hu (1996) extended the Archemedian copula to a class of copulas through mixing max-id copula with the form

$$\begin{aligned} &C_{\psi,H}(u_1, \ldots, u_n) \\ &= \psi\Big(- \sum_{1 \leq i < j \leq n} \log H_{ij}(e^{-p_i \psi^{-1}(u_i)}, e^{-p_i \psi^{-1}(u_j)}) + \sum_{i=1}^{n} v_i p_i \psi^{-1}(u_i)\Big), \end{aligned} \tag{25}$$

where $p_i = (v_i + n - 2)^{-1}, i = 1, \ldots, n$, and v_i is usually non-negative parameter and can be fixed.
An application using the max-id copula in (25), would be in the case of constructing the transition probability of Markov chains with the one-parameter LT functions chosen to be the positive stable and logarithmic series, which are defined as follow. The positive stable function is given by

$$\psi(s) = \exp\{-s^{1/\delta}\}, \quad \delta \geq 1,$$

with corresponding functional inverse given by

$$\psi^{-1}(t) = (-\log t)^{\delta}.$$

The logarithmic series function is given by

$$\psi(s) = -\theta^{-1} \log [1 - (1 - e^{-\delta} e^{-s}], \quad \delta \geq 0,$$

with corresponding functional inverse given by

$$\psi^{-1}(t) = -\log \left[\frac{1 - e^{-\delta t}}{1 - e^{-\delta}} \right].$$

Further detials on such a problem will be discussed in Chapter 4. For further discussions on the copula functions in general and their properties, one can refer to Nelsen (2007) and Joe (2014).

2.3.2 COPULA-BASED COUNT TIME SERIES MODELS

Employing copula to build the correlation structure of continuous time series data has been proposed by many (see for examples, Joe 2014, Guolo and Varin 2014, and Patton 2009). However, for count time series data still not as much literature as in the continuous case due to computational complexity. Joe (2016) suggested copula-based Markov model for time series of counts. Masarotto and Varin (2012) introduced marginal regression models for count time series data with the serial dependence being captured by a Gaussian copula, with a correlation matrix corresponding to a stationary autoregressive moving average (ARMA) process. They preformed statistical inference through an approximated likelihood function

using sequential importance sampling technique. Lennon (2016) and Jia et al. (2018) also applied the same models but suggested different estimation methods including approximate Bayesian computation of the likelihood function and pseudo Gaussian likelihood estimation.

Next, in Chapter 3, we extend the work done in Masarotto and Varin (2012) by including a class with zero-inflated distributions such as the ZIP, ZINB and ZICMP distributions whereas the joint distribution is modeled under Gaussian copula with autoregression moving average (ARMA) errors. In Chapter 4, we introduce a similar class of Markov models to the one in Joe (2016) but with zero-inflated margins. The chapter concentrates on building a class of Markov zero-inflated count time series models based on a joint distribution on consecutive observations. The joint distribution function of the consecutive observations is constructed through copula functions such as the Gaussian, Frank, Gumbel copula functions.

CHAPTER 3

REGRESSION MODEL FOR ZERO-INFLATED COUNT TIME SERIES USING GAUSSIAN COPULA

3.1 INTRODUCTION

Masarotto and Varin (2012) describe Gaussian copula model that underlines the regression setting when covariates are present to model time series data. That is, consider a general regression model with zero-inflated count time series Y_t as response variable, and $\mathbf{X}_t$ as vectors of covariates or independent variables, then the regression model can be represented as:

$$Y_t = g(\mathbf{X}_t, \epsilon_t; \boldsymbol{\theta}), \quad \text{for } t = 1, \ldots, n,$$

where $g(.)$ is a function of the covariates $\mathbf{X}_t$ and the stochastic latent variable or error ϵ_t, which capture the serial dependence. The parameter $\boldsymbol{\theta}$ is a vector of the marginal regression coefficients.

3.2 THE MODEL

We follow the ideas in Masarotto and Varin (2012) and extend them to construct a regression model for zero-inflated time series count data in the presence of covariates. Suppose that the errors ϵ_t for $t = 1, \ldots, n$ follow a stationary $ARMA(p, q)$ process, given in (2), with Gaussian noise, η_t for $t = 1, \ldots, n$ that are independent and identically distributed normal random variables with variance σ_η^2. Then the error vector $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_t)'$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $R(\boldsymbol{\rho})$ where $\boldsymbol{\rho} = (\boldsymbol{\varphi}, \boldsymbol{\delta})$ is a function of the $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)'$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_q)'$, the autoregressive and moving average vector of parameters, respectively. As in Masarotto and Varin (2012), we make the assumption $\sigma_\eta^2 = h(\boldsymbol{\rho})$ so that $R(\boldsymbol{\rho})$ will be a correlation matrix, where $h(\boldsymbol{\rho})$ is given in (4) but with $\gamma(0) = 1$.

Consider as a special case the process $ARMA(1,0)$ (or $AR(1)$). Here the process ϵ_t is governed by

$$\epsilon_t = \varphi\epsilon_{t-1} + \eta_t.$$

With the assumption $\sigma_\eta^2 = 1 - \varphi^2$, the correlation matrix takes the form

$$R(\boldsymbol{\rho}) = R(\boldsymbol{\varphi}) = [\,\varphi^{|i-j|}\,],$$

which is known as autoregressive of order one. Note that the marginally ϵ_t is standard normal and the joint cdf of the vector $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_t)'$ is multivariate normal with mean $\mathbf{0}$ and covariance matrix $R(\boldsymbol{\rho})$. Thus the cdf of $\boldsymbol{\epsilon}$ is $\Phi_{R(\boldsymbol{\rho})}(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ and the induced copula is the Gaussian copula in (23), since $u_t = \Phi(\epsilon_t)$ is uniform on $[0,\ 1]$ for $t = 1, 2, \ldots, n$.

Let F_t be one of the cdfs given in (9), (14) or (20). Following Masarotto and Varin (2012), a general regression model for the zero-inflated count Y_t is

$$Y_t = F_t^{-1}\{\Phi(\epsilon_t)|\mathbf{X}_t; \boldsymbol{\theta}\}, \quad \text{for } t = 1, \ldots, n, \tag{26}$$

where

$$F_t^{-1}(u) = \inf\{z \in \mathbb{R} : F_t(z) \geq u\}, \quad u \in (0,1)$$

is the generalized inverse (quantile function) of the cdf F_t. The vector $\mathbf{X}_t = (\mathbf{x}_t, \mathbf{z}_t, \mathbf{w}_t)'$ consists of covariates corresponding to the intensity (mean) parameter λ_t, the zero-inflation parameter ω_t and the dispersion parameter κ_t if exists, respectively. Notice that some of the covariates could be constant across time. The vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha})'$ is the unknown regression parameter that needs to be estimated from the data.

Constructing the model in (26) in such a way ensures that the zero-inflated count Y_t follows the desired distribution $F_t(.)$ by the integral transformation theorem. Such model appears in the literature under different names (see for examples, Masarotto and Varin, 2012, Jia et al., 2018, and Lennon and Yuan, 2019). Generally, the model falls under the class of nonlinear state-space model since the zero-inflated counts, $\{Y_t\}$ are assumed to be generated using a nonlinear function of the latent or state ARMA process, $\{\epsilon_t\}$.

Another representation of such model is as follow. Assume we know the latent variable ϵ_t, then the zero-inflated count, y_t, is the smallest integer for which the cdf of y_t is greater

than or equal to $\Phi(\epsilon_t)$, i.e. $F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta}) \geq \Phi(\epsilon_t)$, which is proportional to

$$\Phi^{-1}(F_t\{y_t-1|\mathbf{X}_t;\boldsymbol{\theta}\}) < \epsilon_t \leq \Phi^{-1}(F_t\{y_t|\mathbf{X}_t;\boldsymbol{\theta}\}), \quad \text{for } t=1,\ldots,n.$$

That is, the zero inflated count time series are given by

$$Y_t = \begin{cases} 0, & \text{if } \ 0 < u_t \leq F_t\{0|\mathbf{X}_t;\boldsymbol{\theta}\}), \\ y_t, & \text{if } \ F_t\{y_t-1|\mathbf{X}_t;\boldsymbol{\theta}\} < u_t \leq F_t\{y_t|\mathbf{X}_t;\boldsymbol{\theta}\}, \ \ y_t = 1,2,\ldots, \end{cases}$$

where $u_t = \Phi(\epsilon_t) \in (0,1)$, or equivalently

$$Y_t = \begin{cases} 0, & \text{if } \ -\infty < \epsilon_t \leq \Phi^{-1}(F_t\{0|\mathbf{X}_t;\boldsymbol{\theta}\}), \\ y_t, & \text{if } \ \Phi^{-1}(F_t\{y_t-1|\mathbf{X}_t;\boldsymbol{\theta}\}) < \epsilon_t \leq \Phi^{-1}(F_t\{y_t|\mathbf{X}_t;\boldsymbol{\theta}\}), \ \ y_t = 1,2,\ldots, \end{cases}$$

for $t = 1,\ldots,n$.
$F_t\{0|\mathbf{X}_t;\boldsymbol{\theta}\}$ is the probability of $Y_t = 0$ and is given by

$$P(Y_t = 0|\mathbf{X}_t;\boldsymbol{\theta}\}) = \omega_t + (1-\omega_t)e^{-\lambda_t},$$

for the ZIP,

$$P(Y_t = 0|\mathbf{X}_t;\boldsymbol{\theta}\}) = \omega_t + \frac{(1-\omega_t)}{\Gamma(\kappa_t)}\Big(\frac{\kappa_t}{\kappa_t+\lambda_t}\Big)^{\kappa_t}\Gamma(\kappa_t),$$

for the ZINB, and

$$P(Y_t = 0|\mathbf{X}_t;\boldsymbol{\theta}\}) = \omega_t + \frac{(1-\omega_t)}{Z(\lambda_t,\kappa_t)},$$

for the ZICMP.

Note that since the counts are zero-inflated, the probability that the count is zero affects the range of ϵ_t such that the range of u_t when $Y_t = 0$ is wider in comparison with $Y_t > 0$. In other words, the zero-inflation parameter ω_t affects the range of u_t when $Y_t = 0$ whereas the intensity parameter λ_t and the dispersion parameter κ_t (if existed) affect the ranges of u_t when $Y_t > 0$.

The joint distribution function of the zero-inflated count time series, Y_t, for $t = 1,\ldots,n$ follows the Gaussian copula given in (23), that is,

$$F(y_1,\ldots,y_n) \;\; = \;\; P(Y_1 \leq y_1,\ldots,Y_n \leq y_n)$$

$$
\begin{aligned}
&= P(F_1^{-1}\{\Phi(\epsilon_1)|\mathbf{X}_1;\boldsymbol{\theta}\} \leq y_1, \ldots, F_n^{-1}\{\Phi(\epsilon_n)|\mathbf{X}_n;\boldsymbol{\theta}\} \leq y_n) \\
&= P(\Phi(\epsilon_1) \leq F_1(y_1|\mathbf{X}_1;\boldsymbol{\theta}), \ldots, \Phi(\epsilon_n) \leq F_1(y_n|\mathbf{X}_n;\boldsymbol{\theta})) \\
&= P(\epsilon_1 \leq \Phi^{-1}(F_1(y_1|\mathbf{X}_1;\boldsymbol{\theta})), \ldots, \epsilon_n \leq \Phi^{-1}(F_n(y_n|\mathbf{X}_n;\boldsymbol{\theta}))) \\
&= \Phi_{R(\boldsymbol{\rho})}\Big(\Phi^{-1}(F_1(y_1|\mathbf{X}_1;\boldsymbol{\theta})), \ldots, \Phi^{-1}(F_n(y_n|\mathbf{X}_n;\boldsymbol{\theta}))\Big), \qquad (27)
\end{aligned}
$$

and it holds only if (26) holds.

In a linear regression model with normal errors, the correlation of the responses, say Y_t and Y_s, agrees with the correlation of the corresponding errors, ϵ_t and ϵ_s for $t \neq s$. However, in our model the function, F^{-1}, is nonlinear, hence the correlation of Y_t and Y_s is not necessarily linear function of the correlation of ϵ_t and ϵ_s. Jia et al. (2018) studied the relationship between the autocorrelations of the two processes $\{Y_t\}$ and $\{\epsilon_t\}$ and defined a function that links the autocorrelations of the two processes $\{Y_t\}$ and $\{\epsilon_t\}$ using Hermite expansions. Next, we define the function for the ZIP, ZINB, and ZICMP distributions and study the relationship of the autocorrelations of the zero-inflated counts $\{Y_t\}$ and the latent $\{\epsilon_t\}$ process.

3.2.1 AUTOCORRELATION FUNCTION OF THE ZERO-INFLATED COUNTS

In the model (26), the serial dependence or autocorrelation of the observed zero-inflated counts $\{y_t\}$ is captured through the latent process $\{\epsilon_t\}$, which follows an ARMA process with well known autocorrelation structure, which consists of the parameter vector $\boldsymbol{\rho}$ that can be estimated. Hence, to obtain the autocorrelation of the observed zero-inflated count process $\{Y_t\}$, say $\rho_Y(t-s)$, one can define a function that links it to the autocorrelation function of $\{\epsilon_t\}$, say $\rho_\epsilon(t-s)$ for $t \neq s$. Using Hermite expansions (see Chapter 4 of Pipiras and Taqqu, 2017), Jia et al. (2018) derived a function that links the autocovariance function of $\{y_t\}$, i.e. $\gamma_Y(t-s)$ to autocovariance/autocorrelation function of $\{\epsilon_t\}$, i.e. $\gamma_\epsilon(t-s) = \rho_\epsilon(t-s)$ for $t \neq s$, as follow

$$
\gamma_Y(t-s) \;=\; \sum_{k=1}^{\infty} k!\, g_{k,t}(\boldsymbol{\theta}) g_{k,s}(\boldsymbol{\theta}) (\gamma_{\epsilon_t}(t-s))^k
$$

$$= \sum_{k=1}^{\infty} k!\, g_{k,t}(\boldsymbol{\theta}) g_{k,s}(\boldsymbol{\theta})(\rho_\epsilon(t-s))^k, \tag{28}$$

for $t \neq s$ and $t, s = 1, \ldots, n$, where $g_{k,t}(\boldsymbol{\theta})$'s are the Hermite coefficients, and can be given by

$$g_{k,t}(\boldsymbol{\theta}) = \frac{1}{k!} \sum_{m=0}^{\infty} \phi\{\Phi^{-1}(F_t(m|\mathbf{X}_t; \boldsymbol{\theta}))\} H_{k-1}\{\Phi^{-1}(F_t(m|\mathbf{X}_t; \boldsymbol{\theta}))\}, \tag{29}$$

as defined in Jia et al. (2018), where $H_k(.)$ is the Hermite polynomial and is given by

$$H_k(z) = (-1)^k e^{z^2/2} \frac{d^k}{dz^k}\left(e^{-z^2/2}\right), \quad z \in \mathbb{R}.$$

Hence, the autocorrelation function of the zero-inflated counts is given by

$$\begin{aligned} \rho_Y(t-s) &= \frac{\gamma_Y(t-s)}{\sqrt{\gamma_Y(t,t)\gamma_Y(s,s)}} \\ &= \frac{\sum_{k=1}^{\infty} k!\, g_{k,t}(\boldsymbol{\theta}) g_{k,s}(\boldsymbol{\theta})(\rho_\epsilon(t-s))^k}{\sqrt{\gamma_Y(t,t)\gamma_Y(s,s)}}, \end{aligned} \tag{30}$$

where $\gamma_Y(t,t)$ is the variance of Y_t at time t for $t = 1, \ldots, n$, and given in (10), (15), or (21) for the ZIP, ZINB, or ZICMP.

Note that if $\rho_\epsilon(t-s) = 0$, then $\rho_Y(t-s) = 0$, and if $\rho_\epsilon(t-s) = 1$, then

$$\begin{aligned} \rho_Y(t-s) &= \frac{\sum_{k=1}^{\infty} k!\, g_{k,t}(\boldsymbol{\theta}) g_{k,s}(\boldsymbol{\theta})}{\sqrt{\gamma_Y(t,t)\gamma_Y(s,s)}} \\ &= \frac{\sqrt{\gamma_Y(t,t)\gamma_Y(s,s)}}{\sqrt{\gamma_Y(t,t)\gamma_Y(s,s)}} = 1. \end{aligned}$$

However, when $\rho_\epsilon(t-s) = -1$, $\rho_Y(t-s)$ is not necessarily -1. Proof of (28) can be found in page 285, Pipiras and Taqqu (2017).

The following conclusions mentioned in Masarotto and Varin (2012) can be drawn from (30).

1. Given the covariates $\mathbf{X}_t$ and $\mathbf{X}_s$, Y_t and Y_s are independent if ϵ_t and ϵ_s are uncorrelated.

2. Given the covariates $\mathbf{X}_t$ and $\mathbf{X}_s$, the sign of the autocorrelation between Y_t and Y_s is the same as the one corresponding to the autocorrelation between ϵ_t and ϵ_s.

3. Given the covariates $\mathbf{X}_t$ and $\mathbf{X}_s$, the absolute value of the autocorrelation between ϵ_t and ϵ_s is greater or equal than the one corresponding to the autocorrelation between Y_t and Y_s, i.e.
$$|\rho_Y(t-s)|\leq |\rho_\epsilon(t-s)|.$$

These conclusions provide an advantage of using this model over an observation driven model for instance in term of interpretation. In addition, the first property indicates that if the latent process, ϵ_t, follows an $AR(1)$, for example, then the response Y_t only depends on $Y_{t\pm 1}$ for $t = 1, \ldots, n$. The second property is obvious given the non-decreasing function used in the model (26).

Note that when there are no covariates and the zero-inflated count process $\{Y_t\}$ is assumed to be stationary, then (30) becomes

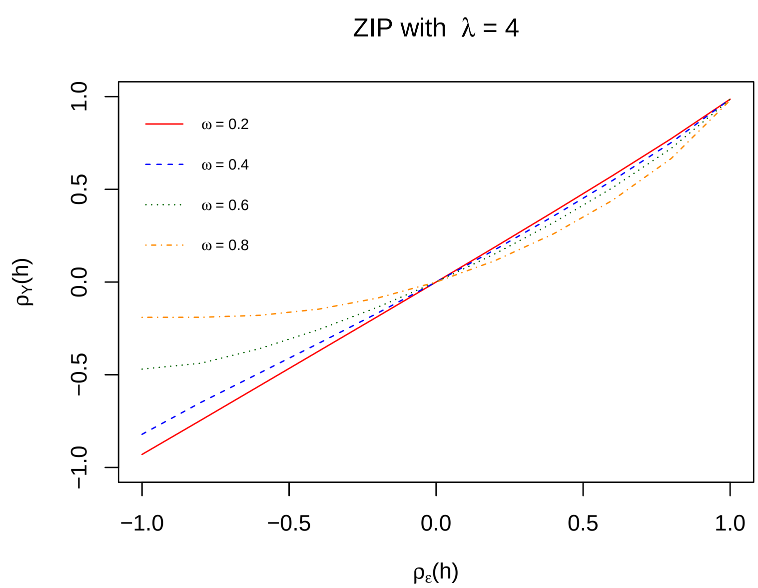
$$\rho_Y(h) = \frac{1}{\gamma_Y(0)} \sum_{k=1}^{\infty} k!\, g_k(\boldsymbol{\theta})^2 (\rho_\epsilon(h))^k \tag{31}$$

for $h = 0, \pm 1, \pm 2, \ldots$ where

$$g_k(\boldsymbol{\theta}) = \frac{1}{k!} \sum_{m=0}^{\infty} \phi\{\Phi^{-1}(F(m;\boldsymbol{\theta}))\} H_{k-1}\{\Phi^{-1}(F(m|\boldsymbol{\theta}))\}.$$

Next, we calculate the Hermite coefficients for the ZIP, ZIINB, and ZICMP marginals, and study the relationships between the autocorrelation of the underlined process $\{\epsilon_t\}$ and the autocorrelation of the observed process $\{Y_t\}$ following one of the ZIP, ZIINB, and ZICMP marginals. First, the expressions in (29) and (30) involve finite summation terms, thus we truncate these terms as suggested in Jia et al. (2018) as follow. For the Hermite coefficients in (29), the distribution function $F_t(m|\mathbf{X}_t;\boldsymbol{\theta})$ converges to one relatively quickly as $m \to \infty$ because we are considering zero-inflated marginal distributions that have light tails. Therefore, the finite summation in (29) is truncated by $m(\boldsymbol{\theta}) - 1$, where $m(\boldsymbol{\theta})$ is the smallest value for which the distribution function $F_t(.)$ is approximately one. For $m \geq m(\boldsymbol{\theta})$, the terms $\phi\{\Phi^{-1}(F_t(m|\mathbf{X}_t;\boldsymbol{\theta}))\}H_{k-1}\{\Phi^{-1}(F_t(m|\mathbf{X}_t;\boldsymbol{\theta}))\}$ in (29) are approximately zero and can be omitted. For the link function in (30), its right hand is multiplied by $\rho_\epsilon(h)^k$, and when $\rho_\epsilon(h)$ is relatively small (which is the case for most

$\{\epsilon_t\}$ considered for the zero-inflated counts), the terms $k!\, g_{k,t}(\boldsymbol{\theta}) g_{k,t+h}(\boldsymbol{\theta}) (\rho_\epsilon(h))^k$ are approximately zeros for large values of k. Thus, truncating the summation in (30) by an appropriate value, say K, will not affect the calculation of the link function.



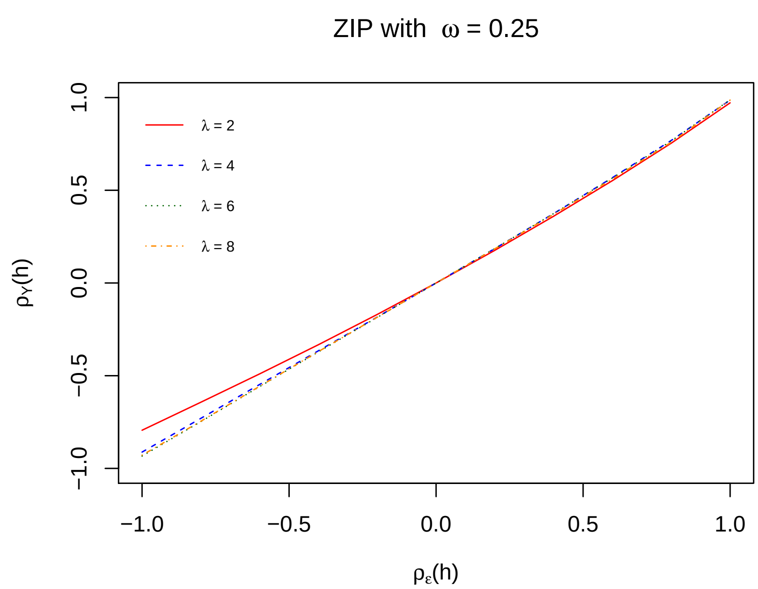


Figure 2: The relationship between $\rho_\epsilon(h)$ and $\rho_Y(h)$ using (31) for the ZIP distribution with $\omega = 0.2, 0.4, 0.6,$ and 0.8 and $\lambda = 4$ (top) and $\omega = 0.25$ and $\lambda = 2, 4, 6,$ and 8 (bottom)

Figure 2 plots the link function in (31) that describes the relationship between the

autocorrelation functions of the state or latent process $\{\epsilon_t\}$ and the zero-inflated count process $\{Y_t\}$ following the ZIP distribution. After truncating the summation in the link function given in (31) by several values of K, we decided to choose $K = 25$ because no significant changes occurred when choosing $K > 25$. The left graph shows the relationship between autocorrelation functions when fixing λ at 4, and changing the values of ω. One can see that as the zero-inflation parameter ω decreases, the line tends to be more linear, i.e. $\rho_\epsilon(h) \approx \rho_Y(h)$, especially when $\rho_\epsilon(h) \in (0, 1)$. However, when $\rho_\epsilon(h) \in (-1, 0)$, the difference between $\rho_\epsilon(h)$ and $\rho_Y(h)$ increases as ω increases, which suggests that negative serial dependence among zero-inflated counts is quite unusual. On the other hand, fixing ω at 0.25 and changing the value of λ as shown in the right graph of Figure 2, one can see minor changes on the line as λ increases. In general, as the mean of the zero-inflated count, Y_t, increases the difference between $\rho_\epsilon(h)$ and $\rho_Y(h)$ decreases.

Figures 3 and 4 show that similar conclusions can be drown when considering the ZINB and ZICMP distribution. The zero-inflated parameter ω plays a significant rule in affecting the relationship between $\rho_\epsilon(h)$ and $\rho_Y(h)$. In addition, when the numerical value of $\rho_Y(h)$ deviates from the theoretical value as in the case when $\rho_\epsilon(h) = 1$, Jia et al. (2018) suggested partial correction of the numerical value of $\rho_Y(h)$. This occurs mostly when the value of $\rho_\epsilon(h)$ is close to one, which is quite rare since we are dealing with low counts (zero-inflated) that usually do not observe strong serial dependence (Joe, 2016).

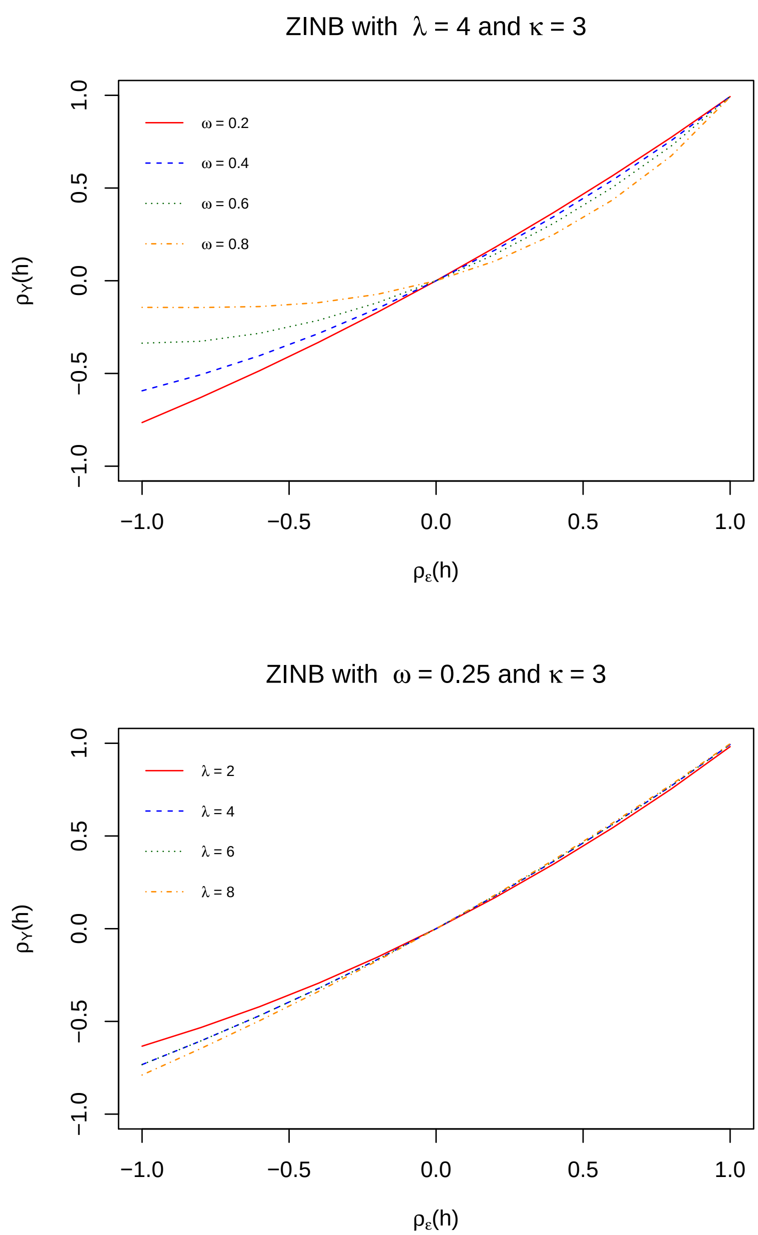
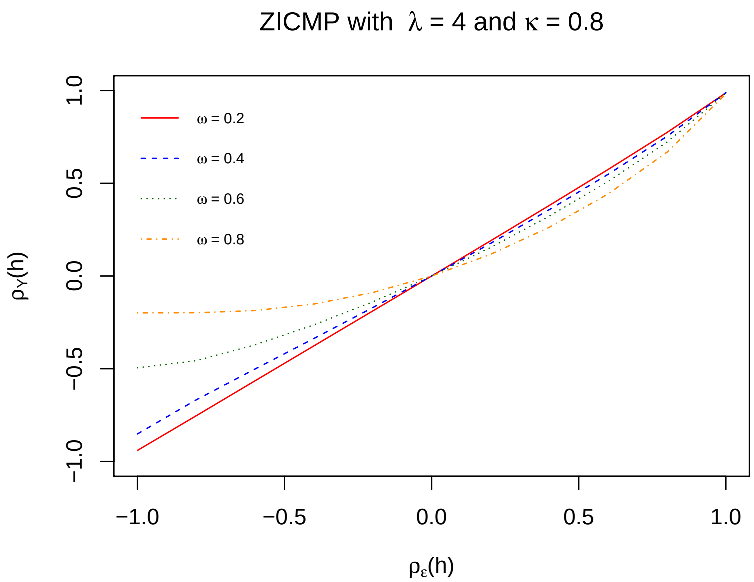


Figure 3: The relationship between $\rho_\epsilon(h)$ and $\rho_Y(h)$ using (31) for the ZINB distribution with $\omega = 0.2, 0.4, 0.6$, and 0.8 and $\lambda = 4$ (top) and $\omega = 0.25$ and $\lambda = 2, 4, 6$, and 8 (bottom). The dispersion parameter $\kappa = 3$ always.



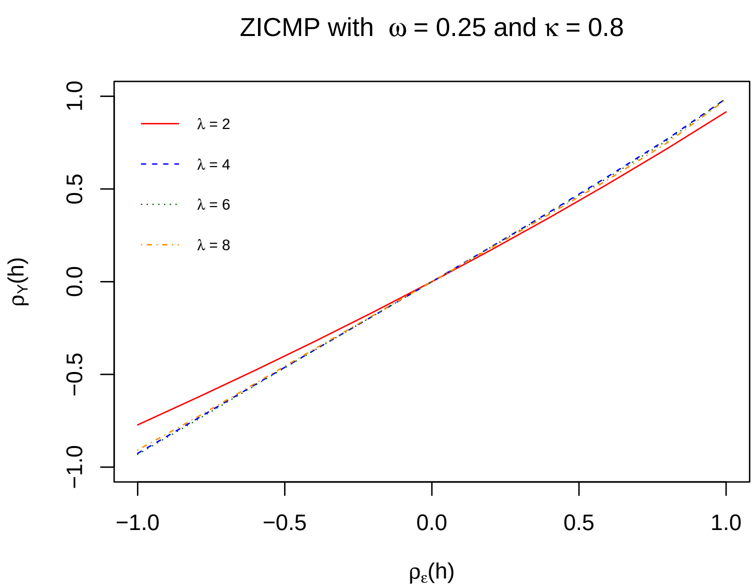


Figure 4: The relationship between $\rho_\epsilon(h)$ and $\rho_Y(h)$ using (31) for the ZICMP distribution with $\omega = 0.2, 0.4, 0.6,$ and 0.8 and $\lambda = 4$ (top) and $\omega = 0.25$ and $\lambda = 2, 4, 6,$ and 8 (bottom). The dispersion parameter $\kappa = 0.9$ always.

3.3 STATISTICAL INFERENCE

3.3.1 PARAMETER ESTIMATION

We intend to estimate the parameter vectors $\boldsymbol{\vartheta} = (\boldsymbol{\theta}, \boldsymbol{\rho})'$ using a maximum likelihood estimation (MLE) method. Based on the probability density function define in (22), the likelihood function is given by

$$\begin{aligned} L(\boldsymbol{\vartheta}; \mathbf{y}) &= \Pr(Y_1 = y_1, \dots, Y_n = y_n) \\ &= \sum_{j_1=0}^{1} \dots \sum_{j_n=0}^{1} (-1)^{j_1+\dots+j_n} F(y_1 - j_1, \dots, y_n - j_n), \end{aligned} \tag{32}$$

where $F(y_1, \dots, y_n)$ for $j_t = 0, 1$ is given in (27), and can be expressed as

$$\Phi_{R(\boldsymbol{\rho})}(\mathcal{D}_1^+, \dots, \mathcal{D}_n^+) = \int_{-\infty}^{\mathcal{D}_1^+} \dots \int_{-\infty}^{\mathcal{D}_n^+} \phi_{R(\boldsymbol{\rho})}(\epsilon_1, \dots, \epsilon_n) d\epsilon_1 \dots d\epsilon_n, \tag{33}$$

where $\mathcal{D}_t^+ = \Phi^{-1}\{F_t(y_t|\mathbf{X}_t; \boldsymbol{\theta})\}$. Therefore, maximizing (32) requires the evaluation of 2^n multivariate distribution functions, and with time series data usually n is quite large so the number of functions will be astronomically large and almost impossible to be optimized. In addition, straightforward optimization methods of the likelihood function are not available yet due to the many-to-one mapping given in (26). In addition, calculating the finite difference in (32) numerically might result in negative values when the dimension is large (Nikoloulopoulos, 2016).

However, for some cases where the copula functions do not have a closed form, the probability density function can be evaluated by integration over a rectangle (Panagiotelis et al., 2012). In fact, for the Gaussian copula with discrete margins, the likelihood function is given by the following n-dimensional rectangular integral

$$L(\boldsymbol{\vartheta}; \mathbf{y}) = \Pr(Y_1 = y_1, \dots, Y_n = y_n)$$

$$= \int\limits_{\mathcal{D}_1(y_1;\boldsymbol{\theta})} \cdots \int\limits_{\mathcal{D}_n(y_n;\boldsymbol{\theta})} \phi_{R(\boldsymbol{\rho})}(\epsilon_1, \ldots, \epsilon_n) d\epsilon_1 \ldots d\epsilon_n,$$

$$= \frac{1}{\sqrt{|R(\boldsymbol{\rho})|(2\pi)^n}} \int\limits_{\mathcal{D}_1(y_1;\boldsymbol{\theta})} \cdots \int\limits_{\mathcal{D}_n(y_n;\boldsymbol{\theta})} e^{-\frac{1}{2}\boldsymbol{\epsilon}' R(\boldsymbol{\rho})^{-1}\boldsymbol{\epsilon}} d\boldsymbol{\epsilon} \tag{34}$$

where

$$\mathcal{D}_t(y_t;\boldsymbol{\theta}) = [\Phi^{-1}\{F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta})\}, \Phi^{-1}\{F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta})\}] \tag{35}$$

for $t = 1, \ldots, n$ and $\phi_{R(\boldsymbol{\rho})}(.)$ is the probability density function of an n-dimensional normal distribution with zero mean vector and a variance covariance matrix given by $R(\boldsymbol{\rho})$. For small n, notable works have been done on precisely approximating the normal integral given in (34) (see for examples, Joe 1995 and Genz 1992). However, for large n, as of the case for time series data, evaluating the likelihood function using these deterministic approximations is computationally intensive and is inefficient especially when the number of covariates is large.

Recently, several techniques emerged to estimate copula based multivariate models with large dimension. Some of these techniques employ Monte Carlo approximation methods to obtain the ML estimates of the parameter such as the Monte Carlo EM algorithm by Lennon (2016) and the sequential importance sampling by Masarotto and Varin (2012) and Jia et al. (2018). Jia et al. (2018) also suggested applying pseudo Gaussian likelihood estimation method that is both simpler than the sequential importance sampling method and comparable to in terms of efficiency. Bayesian estimation methods were also considered for estimating Gaussian copula models in Pitt et al. (2006). Panagiotelis et al. (2012) extended the the principles of vine pair copula constructions to discrete margins, which significantly reduced the computational burden of evaluating the 2^n multivariate copulas to just evaluating $2n(n-1)$ bivariate copula functions. Lennon (2016) applied the pair copula construction technique to time series of count with negative binomial margins. Here we obtain the ML estimates of our model's parameters using sequential importance sampling that was suggested by Masarotto and Varin (2012). In the next subsection, we describe the sequential importance sampling method used to approximate the likelihood function for the zero-inflated count responses and discuss in details some special cases.

3.3.2 SEQUENTIAL IMPORTANCE SAMPLING

Masarotto and Varin (2012) argued that applying simple Monte Carlo approximations of the likelihood given in (34) used in importance sampling (IS) are quite inefficient. However, they suggested sequential importance sampling method inspired by the popular Geweke-Hajivassiliou-Keane (GHK) algorithm (Geweke, 1991; Hajivassiliou et al., 1996; Keane, 1994) which was proven to be quite efficient in approximating the multivariate probability integral given in (34). They assumed sampling from the following truncated normal density given by

$$f_t(\epsilon_t|y_t, \epsilon_{t-1}, \ldots, \epsilon_1; \boldsymbol{\rho}), \quad t = 1, \ldots, n \tag{36}$$

as a replacement of the difficult to control, $f_t(\epsilon_t|y_t, y_{t-1}, \ldots, y_1; \boldsymbol{\rho})$ over the interval given in (35). In addition, since we assume that the joint distribution of the errors is multivariate normal distribution with variance covariance matrix $R(\boldsymbol{\rho})$, the conditional density $\phi(\epsilon_t|\epsilon_{t-1}, \ldots, \epsilon_1; \boldsymbol{\rho})$ is of univariate normal distribution with mean

$$m_t = \mathrm{E}(\epsilon_t|\epsilon_{t-1}, \ldots, \epsilon_1) \tag{37}$$

and variance

$$v_t^2 = \mathrm{Var}(\epsilon_t|\epsilon_{t-1}, \ldots, \epsilon_1), \tag{38}$$

for $t = 1, \ldots, n$. The quantities m_t and v_t^2 can be efficiently obtained through the Cholesky decomposition of $R(\boldsymbol{\rho})$. That is, letting $\boldsymbol{\epsilon} = L\mathbf{z}$ where $\mathbf{z} \sim N_n(\mathbf{0}, I_n)$, in which LL' is the Cholesky decomposition of the covariance matrix $R(\boldsymbol{\rho})$ since it is assumed that $R(\boldsymbol{\rho})$ here is a symmetric positive definite matrix, i.e. $R(\boldsymbol{\rho}) = LL'$. In addition, L is a lower triangular matrix with real and positive diagonal elements, i.e.

$$L = \begin{bmatrix} l_{11} & 0 & 0 & \ldots & 0 & 0 \\ l_{21} & l_{22} & 0 & \ldots & 0 & 0 \\ l_{31} & l_{32} & l_{33} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ l_{n-11} & l_{n-12} & l_{n-13} & \ldots & l_{n-1n-1} & 0 \\ l_{n1} & l_{n2} & l_{n3} & \ldots & l_{nn-1} & l_{nn} \end{bmatrix}_{n\times n},$$

with the components are generally given by

$$l_{tt} = \Big(r_{tt} - \sum_{j=1}^{t-1} l_{tj}^2\Big)^{1/2}, \quad \forall t, \tag{39}$$

and

$$l_{it} = \frac{1}{l_{tt}}\Big(r_{it} - \sum_{j=1}^{t-1} l_{ij}l_{tj}\Big), \quad \forall i > t, \tag{40}$$

where r_{it} are the components of the correlation matrix $R(\boldsymbol{\rho})$. Also,

$$\boldsymbol{\epsilon}'R(\boldsymbol{\rho})^{-1}\boldsymbol{\epsilon} = \mathbf{z}'L'(LL')^{-1}L\mathbf{z} = \mathbf{z}'L'(L')^{-1}L^{-1}L\mathbf{z} = \mathbf{z}'\mathbf{z},$$

and

$$d\boldsymbol{\epsilon} = dL\mathbf{z} = |L|d\mathbf{z} = |R(\boldsymbol{\rho})|^{1/2}d\mathbf{z}.$$

Hence, if $\boldsymbol{D}^- \leq \boldsymbol{\epsilon} \leq \boldsymbol{D}^+$, where

$$\boldsymbol{D}^- = (\Phi^{-1}\{F_1(y_1^-|\mathbf{X}_t;\boldsymbol{\theta})\}, \ldots, \Phi^{-1}\{F_n(y_n^-|\mathbf{X}_t;\boldsymbol{\theta})\})',$$

and

$$\boldsymbol{D}^+ = (\Phi^{-1}\{F_1(y_1|\mathbf{X}_t;\boldsymbol{\theta})\}, \ldots, \Phi^{-1}\{F_n(y_n|\mathbf{X}_t;\boldsymbol{\theta})\})',$$

then $\boldsymbol{D}^- \leq L\mathbf{z} \leq \boldsymbol{D}^+ \iff L^{-1}\boldsymbol{D}^- \leq \mathbf{z} \leq L^{-1}\boldsymbol{D}^+$. That is,

$$\begin{bmatrix} \frac{\mathcal{D}_1^-}{l_{11}} \\ \frac{\mathcal{D}_2^- - l_{21}z_1}{l_{22}} \\ \vdots \\ \frac{\mathcal{D}_n^- - \sum_{t=1}^{n-1} l_{nt}z_t}{l_{nn}} \end{bmatrix} \leq \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \leq \begin{bmatrix} \frac{\mathcal{D}_1^+}{l_{11}} \\ \frac{\mathcal{D}_2^+ - l_{21}z_1}{l_{22}} \\ \vdots \\ \frac{\mathcal{D}_n^+ - \sum_{t=1}^{n-1} l_{nt}z_t}{l_{nn}} \end{bmatrix}$$

Thus, following Genz (1992)'s transformation method, the multivariate integral given in (34) can be expressed as:

$$L(\boldsymbol{\vartheta};\mathbf{y}) \;=\; \frac{1}{\sqrt{(2\pi)^n}} \int_{a_1'}^{b_1'} e^{-\frac{z_1^2}{2}} \int_{a_2'(z_1)}^{b_2'(z_1)} e^{-\frac{z_2^2}{2}} \ldots \int_{a_n'(z_1,\ldots,z_{n-1})}^{b_n'(z_1,\ldots,z_{n-1})} e^{-\frac{z_n^2}{2}} d\mathbf{z},$$

with

$$a'_t(z_1, \ldots, z_{t-1}) = \frac{\mathcal{D}_t^- - \sum_{j=1}^{t-1} l_{tj} z_j}{l_{tt}}$$

and

$$b'_t(z_1, \ldots, z_{t-1}) = \frac{\mathcal{D}_t^+ - \sum_{j=1}^{t-1} l_{tj} z_j}{l_{tt}}$$

Now, we can transform the z_t's separately using Masarotto and Varin (2012)'s idea by setting $z_t = \epsilon_t(u_t)$ where $\epsilon_t(u_t)$ is given by

$$\epsilon_t(u_t) = m_t + v_t \Phi^{-1}\{(1-u_t)a_t + u_t b_t\}, \quad t = 1, \ldots, n \tag{41}$$

to draw from the truncated normal density given in (36) where $u_1, \ldots, u_n$ are n independent draws from a uniform random variable on the unit interval $(0, 1)$,

$$a_t = \Phi\left[\frac{\Phi^{-1}\{F_t(y_t^-|\mathbf{X}_t; \boldsymbol{\theta})\} - m_t}{v_t}\right], \tag{42}$$

and

$$b_t = \Phi\left[\frac{\Phi^{-1}\{F_t(y_t|\mathbf{X}_t; \boldsymbol{\theta})\} - m_t}{v_t}\right], \tag{43}$$

for $t = 1, \ldots, n$.

The likelihood function is then approximated by the following sequential sampler algorithm.

1. For $k = 1, \ldots, K$,
   (a) generate n independent uniform$(0, 1)$ random variables, $u_1^{(k)}, \ldots, u_n^{(k)}$;
   (b) compute the randomized errors $\epsilon_t^{(k)} = \epsilon_t(u_t^{(k)})$ using (41);
2. estimate the likelihood by:

$$\widehat{L}(\boldsymbol{\vartheta}; \mathbf{y}) = \frac{1}{K} \sum_{k=1}^{K} \left\{ \prod_{t=1}^{n} \frac{\phi(\epsilon_t^{(k)}|\epsilon_{t-1}^{(k)}, \ldots, \epsilon_1^{(k)}; \boldsymbol{\vartheta})}{f_t(\epsilon_t^{(k)}|y_t, \epsilon_{t-1}^{(k)}, \ldots, \epsilon_1^{(k)}; \boldsymbol{\vartheta})} \right\}, \tag{44}$$

where K denotes the number of replication. Börsch-Supan and Hajivassiliou (1993) give a way to show that $\widehat{L}(\boldsymbol{\vartheta}; \mathbf{y})$ is an unbiased estimator of $L(\boldsymbol{\vartheta}; \mathbf{y})$. The following lemma is

similar to the one in their paper.

Lemma 1. *The likelihood approximation $\widehat{L}(\boldsymbol{\vartheta}; \mathbf{y})$ given in (44) is an unbiased estimator of $L(\boldsymbol{\vartheta}; \mathbf{y})$.*

Proof. It is sufficient to show the Lemma for $K = 1$. The expected value of $\widehat{L}(\boldsymbol{\vartheta}; \mathbf{y})$ is given by

$$E[\widehat{L}] = \int \widehat{L}(\mathbf{z}) f(\mathbf{z}) d\mathbf{z},$$

where $f(\mathbf{z})$ denotes the sequential truncated draws that generate (36), i.e.

$$\begin{aligned} f(\mathbf{z}) &= \prod_{t=1}^{n} f_t(z_t | y_t, z_{t-1}, \ldots, z_1) \qquad \text{if} \qquad T(\mathbf{y}) = \boldsymbol{D}^- \leq L\mathbf{z} \leq \boldsymbol{D}^+ \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise.} \end{aligned}$$

By the definition of $\widehat{L}$, (44),

$$\begin{aligned} E[\widehat{L}] &= \int_{-\infty}^{\infty} \left(\prod_{t=1}^{n} f_t(z_t | y_t, z_{t-1}, \ldots, z_1) \right) \\ &\times \left(\prod_{t=1}^{n} \frac{\phi(z_t | z_{t-1}, \ldots, z_1)}{f_t(z_t | y_t, z_{t-1}, \ldots, z_1)} \right) dz_1 \ldots dz_n \\ &= \int_{-\infty}^{\infty} \prod_{t=1}^{n} \phi(z_t | z_{t-1}, \ldots, z_1) d\mathbf{z} \\ &= \int_{T(\mathbf{y})} \prod_{t=1}^{n} \phi(z_t | z_{t-1}, \ldots, z_1) d\mathbf{z} \\ &= \Pr(\boldsymbol{D}^- \leq L\mathbf{z} \leq \boldsymbol{D}^+) = L(\boldsymbol{\theta}, \boldsymbol{\rho}; \mathbf{y}). \end{aligned}$$

Therefore, $\widehat{L}(\boldsymbol{\vartheta}; \mathbf{y})$ is an unbiased estimator of $L(\boldsymbol{\vartheta}; \mathbf{y})$. □

Thus, the maximum likelihood estimate of $\boldsymbol{\vartheta}$ can be obtained by:

$$\widehat{\boldsymbol{\vartheta}} = \arg\max_{\boldsymbol{\vartheta}} \widehat{L}(\boldsymbol{\vartheta}; \mathbf{y}). \tag{45}$$

This optimization will yield a Hessian Matrix that can be inverted to obtain standard

errors for the model parameters. Other possibilities of calculating the standard errors for the model parameters are through the heteroscedasticity and autocorrelation consistent sandwich estimators of the standard errors (Andrews, 1991) and a block bootstrapping (Lahiri, 2013). The latter allows for deriving confidence intervals too, and under (26), a block bootstrap of the process $\{Y_t\}$ corresponds to that of $\{\epsilon_t\}$ (Jia et al., 2018).

Special Cases: $AR(1)$ and $MA(1)$ Correlation Structures

As a special case of the proceeding algorithm, assume the latent process $\{\epsilon_t\}$ follows an $AR(1)$ process. The correlation matrix, $R(\boldsymbol{\rho})$, is then given by

$$R(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \varphi & \varphi^2 & \dots & \varphi^{n-2} & \varphi^{n-1} \\ \varphi & 1 & \varphi & \dots & \varphi^{n-3} & \varphi^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \varphi^{n-2} & \varphi^{n-3} & \varphi^{n-4} & \dots & 1 & \varphi \\ \varphi^{n-1} & \varphi^{n-2} & \varphi^{n-3} & \dots & \varphi & 1 \end{bmatrix}_{n\times n} .$$

The Cholesky factorization of $R(\boldsymbol{\rho})$ is LL', and the lower triangular components of L can be found using (39) and (40). In particular, the components of first column of L are given by

$$l_{11} = 1,$$

$$l_{21} = \varphi,$$

$$l_{31} = \varphi^2,$$

$$\vdots$$

$$l_{i1} = \varphi^{i-1},$$

$$\vdots$$

$$l_{n-11} = \varphi^{n-2},$$

$$l_{n1} = \varphi^{n-1},$$

the components of second column are

$$l_{22} = (1 - \varphi^2)^{1/2},$$

$$l_{32} = \frac{1}{(1-\varphi^2)^{1/2}}(\varphi - \varphi^3) = \varphi(1-\varphi^2)^{1/2},$$

$$\vdots$$

$$l_{i2} = \frac{1}{(1-\varphi^2)^{1/2}}(\varphi^{i-2} - \varphi^{i-1}\varphi) = \varphi^{i-2}(1-\varphi^2)^{1/2},$$

$$\vdots$$

$$l_{n-12} = \varphi^{n-3}(1-\varphi^2)^{1/2},$$

$$l_{n2} = \varphi^{n-2}(1-\varphi^2)^{1/2},$$

and the components of third column are

$$l_{33} = \{1 - [\varphi^4 + \varphi^2(1-\varphi^2)]\}^{1/2} = (1-\varphi^2)^{1/2},$$

$$\vdots$$

$$\begin{aligned} l_{i3} &= \frac{1}{(1-\varphi^2)^{1/2}}[\varphi^{i-3} - (\varphi^{i+1} + (1-\varphi^2)\varphi^{i-1}] \\ &= \frac{1}{(1-\varphi^2)^{1/2}}[\varphi^{i-3} - \varphi^{i+1} - \varphi^{i-1} + \varphi^{i+1}] \\ &= \frac{\varphi^{i-3}}{(1-\varphi^2)^{1/2}}(1-\varphi^2) = \varphi^{i-3}(1-\varphi^2)^{1/2}, \end{aligned}$$

$$\vdots$$

$$l_{n-13} = \varphi^{n-4}(1-\varphi^2)^{1/2},$$

$$l_{n3} = \varphi^{n-3}(1-\varphi^2)^{1/2}.$$

Following the same pattern, a general form of the below diagonal components of L for $i > t > 1$ is given by

$$l_{it} = \varphi^{i-t}(1-\varphi^2)^{1/2}.$$

and a general form of the diagonals of L for $t > 1$ is given by

$$\begin{aligned}
l_{tt} &= \left\{1 - \left[\varphi^{2(t-1)} + (1-\varphi^2)\varphi^{2(t-2)} + \cdots + (1-\varphi^2)\varphi^{2(2)} + (1-\varphi^2)\varphi^2\right]\right\}^{1/2} \\
&= \left\{1 - \left[\varphi^{2(t-1)} + (1-\varphi^2)\sum_{j=2}^{t-1}\varphi^{2(j-1)}\right]\right\}^{1/2} \\
&= \left\{1 - \left[\varphi^{2(t-1)} - (1-\varphi^2)\varphi^{2(t-1)} + (1-\varphi^2)\sum_{j=1}^{t-1}\varphi^{2(j-1)}\right]\right\}^{1/2} \\
&= \left\{1 - \left[\varphi^{2t} + (1-\varphi^2)\sum_{j=1}^{t-1}\varphi^{2(j-1)}\right]\right\}^{1/2} \\
&= \left\{1 - \left[\varphi^{2t} + (1-\varphi^2)\frac{\varphi^2 - \varphi^{2t}}{1-\varphi^2}\right]\right\}^{1/2} \\
&= \left\{1 - \left[\varphi^{2t} + \varphi^2 - \varphi^{2t}\right]\right\}^{1/2} \\
&= (1-\varphi^2)^{1/2}.
\end{aligned}$$

In summary,

$$l_{tt} = \begin{cases} 1 & \text{for } \ t = 1, \\ (1-\varphi^2)^{1/2} & \text{for } \ t > 1, \end{cases}$$

and

$$l_{it} = \begin{cases} \varphi^{i-1} & \text{for } \ t = 1, \\ \varphi^{i-t}(1-\varphi^2)^{1/2} & \text{for } \ i > t > 1, \end{cases}$$

so that

$$L = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \varphi & (1-\varphi^2)^{1/2} & 0 & \dots & 0 & 0 \\ \varphi^2 & \varphi(1-\varphi^2)^{1/2} & (1-\varphi^2)^{1/2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \varphi^{n-2} & \varphi^{n-3}(1-\varphi^2)^{1/2} & \varphi^{n-4}(1-\varphi^2)^{1/2} & \dots & (1-\varphi^2)^{1/2} & 0 \\ \varphi^{n-1} & \varphi^{n-2}(1-\varphi^2)^{1/2} & \varphi^{n-3}(1-\varphi^2)^{1/2} & \dots & \varphi(1-\varphi^2)^{1/2} & (1-\varphi^2)^{1/2} \end{bmatrix}_{n\times n}.$$

Hence, the conditional mean and variance given in (37) and (38), respectively, are given by

$$\begin{aligned} m_t &= \sum_{j=1}^{t-1} l_{tj}\epsilon_j \\ &= \varphi^{t-1}\epsilon_1 + \varphi^{t-2}(1-\varphi^2)^{1/2}\epsilon_2 + \cdots + \varphi(1-\varphi^2)^{1/2}\epsilon_{t-1} \\ &= \varphi^{t-1}\epsilon_1 + (1-\varphi^2)^{1/2}\sum_{j=2}^{t-1}\varphi^{t-j}\epsilon_j \end{aligned}$$

for $t > 1$ and $m_t = 0$ for $t = 1$, and

$$v_t^2 = l_{tt}^2 = (1-\varphi^2)$$

for $t > 1$ and $v_t^2 = 1$ for $t = 1$.Thus, the quantities defined in (42) and (43) are given by

$$a_t = \Phi\left[\frac{\Phi^{-1}\{F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta})\} - \left(\varphi^{t-1}\epsilon_1 + (1-\varphi^2)^{1/2}\sum_{j=2}^{t-1}\varphi^{t-j}\epsilon_j\right)}{(1-\varphi^2)^{1/2}}\right],$$

for $t > 1$ and $a_t = F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta})$ for $t = 1$, and

$$b_t = \Phi\left[\frac{\Phi^{-1}\{F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta})\} - \left(\varphi^{t-1}\epsilon_1 + (1-\varphi^2)^{1/2}\sum_{j=2}^{t-1}\varphi^{t-j}\epsilon_j\right)}{(1-\varphi^2)^{1/2}}\right],$$

for $t > 1$ and $b_t = F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta})$ for $t = 1$, respectively.

Next, assume the latent process $\{\epsilon_t\}$ follows an $MA(1)$ process. The correlation matrix,

$R(\boldsymbol{\rho})$, is then given by

$$R(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \delta_1 & 0 & \dots & 0 & 0 \\ \delta_1 & 1 & \delta_1 & \dots & 0 & 0 \\ 0 & \delta_1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & \delta_1 \\ 0 & 0 & 0 & \dots & \delta_1 & 1 \end{bmatrix}_{n\times n},$$

where $\delta_1 = \delta/(1+\delta^2)$. The Cholesky factorization of $R(\boldsymbol{\rho})$ is LL', and the lower triangular components of L can be found using (39) and (40), which reduce to be given by

$$l_{tt} = \begin{cases} 1 & \text{for} \quad t = 1, \\ (1 - l_{tt-1})^{1/2} & \text{for} \quad t > 1, \end{cases}$$

and

$$l_{it} = \begin{cases} \frac{r_{ii-1}}{l_{i-1i-1}} & \text{for} \quad t = i-1, \\ 0 & \text{for} \quad t < i-1, \end{cases}$$

which can be calculated recursively. Hence, the lower triangular matrix L has the form

$$L = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ r_{21} & (1-l_{21})^2 & 0 & \dots & 0 & 0 \\ 0 & \frac{r_{32}}{l_{22}} & (1-l_{32})^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & (1-l_{n-1n-2})^{1/2} & 0 \\ 0 & 0 & 0 & \dots & \frac{r_{nn-1}}{l_{n-1n-1}} & (1-l_{nn-1})^{1/2} \end{bmatrix}_{n\times n}.$$

To obtain expressions of l_{tt} and l_{it}, consider the first several terms, that is,

$$
\begin{aligned}
l_{11} &= 1, & l_{21} &= \delta_1, \\
l_{22} &= (1-\delta_1^2)^{1/2}, & l_{32} &= \delta_1\left(\frac{1}{1-\delta_1^2}\right)^{1/2}, \\
l_{33} &= \left(\frac{1-2\delta_1^2}{1-\delta_1^2}\right)^{1/2}, & l_{34} &= \delta_1\left(\frac{1-\delta_1^2}{1-2\delta_1^2}\right)^{1/2}, \\
l_{44} &= \left(\frac{1-3\delta_1^2+\delta_1^4}{1-2\delta_1^2}\right)^{1/2}, & l_{54} &= \delta_1\left(\frac{1-2\delta_1^2}{1-3\delta_1^2+\delta_1^4}\right)^{1/2}, \\
l_{55} &= \left(\frac{1-4\delta_1^2+3\delta_1^4}{1-3\delta_1^2+\delta_1^4}\right)^{1/2}, & l_{65} &= \delta_1\left(\frac{1-3\delta_1^2+\delta_1^4}{1-4\delta_1^2+3\delta_1^4}\right)^{1/2}, \\
l_{66} &= \left(\frac{1-5\delta_1^2+6\delta_1^4-\delta_1^6}{1-4\delta_1^2+3\delta_1^4}\right)^{1/2}, & l_{76} &= \delta_1\left(\frac{1-4\delta_1^2+3\delta_1^4}{1-5\delta_1^2+6\delta_1^4-\delta_1^6}\right)^{1/2}, \\
l_{77} &= \left(\frac{1-6\delta_1^2+10\delta_1^4-4\delta_1^6}{1-5\delta_1^2+6\delta_1^4-\delta_1^6}\right)^{1/2}, & l_{87} &= \delta_1\left(\frac{1-5\delta_1^2+6\delta_1^4-\delta_1^6}{1-6\delta_1^2+10\delta_1^4-4\delta_1^6}\right)^{1/2}.
\end{aligned}
$$

Following the pattern, we find that

$$
l_{tt} = \begin{cases} G_{1t}^{1/2} & \text{if } t \text{ is even,} \\ G_{2t}^{1/2} & \text{if } t \text{ is odd,} \end{cases} \tag{46}
$$

where,

$$
G_{1t} = \frac{\sum_{j=0}^{t/2}(-1)^j\binom{t-j}{j}\delta_1^{2j}}{\sum_{j=0}^{t/2-1}(-1)^j\binom{t-j-1}{j}\delta_1^{2j}},
$$

and

$$
G_{2t} = \frac{\sum_{j=0}^{(t-1)/2}(-1)^j\binom{t-j}{j}\delta_1^{2j}}{\sum_{j=0}^{(t-1)/2}(-1)^j\binom{t-j-1}{j}\delta_1^{2j}}.
$$

Then,

$$
l_{it} = \begin{cases} \delta_1 G_{1t}^{-1/2} & \text{if } t \text{ is even,} \\ \delta_1 G_{2t}^{-1/2} & \text{if } t \text{ is odd,} \end{cases} \tag{47}
$$

for $i > 1$. Hence, the conditional mean and variance given in (37) and (38), respectively, are

$$
m_t = \begin{cases} \delta_1 G_{2t}^{-1/2}\epsilon_{t-1} & \text{if } t \text{ is even,} \\ \delta_1 G_{1t}^{-1/2}\epsilon_{t-1} & \text{if } t \text{ is odd,} \end{cases}
$$

for $t > 1$ and $m_t = 0$ for $t = 1$, and

$$v_t^2 = \begin{cases} G_{1t} & \text{if } t \text{ is even,} \\ G_{2t} & \text{if } t \text{ is odd,} \end{cases}$$

Thus, the quantities defined in (42) and (43) are given by

$$a_t = \begin{cases} \Phi\left[\frac{\Phi^{-1}\{F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta})\}-\delta_1 G_{2t}^{-1/2}\epsilon_t}{G_{12}^{1/2}}\right] & \text{if } t \text{ is even,} \\ \\ \Phi\left[\frac{\Phi^{-1}\{F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta})\}-\delta_1 G_{1t}^{-1/2}\epsilon_t}{G_{1t}^{1/2}}\right] & \text{if } t \text{ is odd,} \end{cases}$$

for $t > 1$ and $a_t = F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta})$ for $t = 1$, and

$$b_t = \begin{cases} \Phi\left[\frac{\Phi^{-1}\{F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta})\}-\delta_1 G_{2t}^{-1/2}\epsilon_t}{G_{12}^{1/2}}\right] & \text{if } t \text{ is even,} \\ \\ \Phi\left[\frac{\Phi^{-1}\{F_t(y_t-|\mathbf{X}_t;\boldsymbol{\theta})\}-\delta_1 G_{1t}^{-1/2}\epsilon_t}{G_{1t}^{1/2}}\right] & \text{if } t \text{ is odd,} \end{cases}$$

for $t > 1$ and $b_t = F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta})$ for $t = 1$, respectively.
For higher order dependence structures, one can numerically compute the Cholesky decomposition matrices.

3.4 MODEL ASSESSMENT

3.4.1 RESIDUALS

To check the goodness of a regression model fit, residuals analysis is often the first choice to consider. For an ordinary linear regression model with independent normal responses, considerable literature and techniques were developed to analyze the residuals. Most of these techniques suggest obtaining normally distributed residuals, which indicate the regression model is adequately fitted. For non-normal responses, discrete specifically, using the usual residuals, that is the difference between the fitted and the predicted values, provides residuals that might depart from normality. Dunn and Smyth (1996) introduced

randomized quantile residuals that are i.i.d. as standard normal even if the responses are discrete and dependent. These residuals are given by

$$r_t(u_t) = \Phi^{-1}\{q_t^- + u_t(q_t - q_t^-)\}, \quad t = 1, \dots, n, \tag{48}$$

where $q_t = F_t(y_t|y_{t-1}, \dots, y_1; \widehat{\boldsymbol{\vartheta}})$, $q_t^- = F_t(y_t^-|y_{t-1}, \dots, y_1; \widehat{\boldsymbol{\vartheta}})$, and u_t's are draw from the uniform $(0, 1)$ distribution. The randomization component, u_t is to insure the residuals are independent and continuous. Assuming the assumptions of our model holds, these residuals will be normally distributed (Masarotto and Varin, 2012). Since there is a randomization component in (48), it is advised to plot the quantile residuals multiple times and see if there is any consistent pattern, otherwise it should be ignored (Dunn and Smyth, 1996).

3.4.2 PREDICTION

An important advantage of using the model (26) is that prediction of the time series of counts, Y_t, can be obtained directly once we predict the latent ARMA process, $\{\epsilon_t\}$, given in Equation (2). Jia et al. (2018) suggested one way to estimate the latent process by the conditional expectation of ϵ_t given the observed zero-inflated count Y_t, which is nothing but the mean of the truncated normal distribution over the interval $(\Phi^{-1}\{F_1(y_1^-; \boldsymbol{\theta})\}, \Phi^{-1}\{F_1(y_1; \boldsymbol{\theta})\}]$ (see Appendix A.1 for more details). That is, ϵ_t is estimated by

$$\widehat{\epsilon}_t = \mathrm{E}(\epsilon_t|Y_t = y_t) = \frac{\phi(\Phi^{-1}\{F_t(y_t^-|\mathbf{X}_t; \boldsymbol{\theta})\}) - \phi(\Phi^{-1}\{F_t(y_t|\mathbf{X}_t; \boldsymbol{\theta})\})}{F_t(y_t|\mathbf{X}_t; \boldsymbol{\theta}) - F_t(y_t^-|\mathbf{X}_t; \boldsymbol{\theta})}, \tag{49}$$

for $t = 1, \dots, n$. Hence, one can apply standard prediction methods of ARMA process to predict ϵ_{n+1}. For example, a one-step-ahead prediction of $AR(1)$ process using best linear prediction method is given by

$$\tilde{\epsilon}_{n+1} = \widehat{\varphi}\, \widehat{\epsilon}_n,$$

where $\widehat{\varphi}$ is the ML estimate of the autocorrelation of the latent process. See Shumway and Stoffer (2011) for more details on predicting ARMA process. In application, the predicted value of Y_t is given by

$$\tilde{Y}_{n+1} = F_{n+1}^{-1}\{\Phi(\tilde{\epsilon}_{n+1})|\mathbf{X}_{n+1}; \widehat{\boldsymbol{\theta}}\}, \tag{50}$$

where $\widehat{\boldsymbol{\theta}}$ is the ML estimate of the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$, and $\tilde{\epsilon}_{n+1}$ is a one-step-ahead

prediction for ϵ_{n+1}.

Another way of prediction, is to consider the conditional expectation of Y_t given the past $Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1$, which is given by

$$\begin{aligned} &\mathrm{E}(Y_t | Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1; \boldsymbol{\vartheta}) = \\ &\sum_{y_t \in \mathcal{S}} y_t \mathrm{Pr}(Y_t = y_t | Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1; \boldsymbol{\vartheta}) = \\ &\sum_{y_t \in \mathcal{S}} y_t \frac{\mathrm{Pr}(Y_t = y_t, Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1; \boldsymbol{\vartheta})}{\mathrm{Pr}(Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1; \boldsymbol{\vartheta})}. \end{aligned} \tag{51}$$

3.5 SIMULATED EXAMPLES

To evaluate the performance of the proposed method, comprehensive simulation studies are conducted. We carry out the simulation in the statistical software R (R Core Team, 2013). In Subsection 3.5.1, we provide simple simulated examples in order to understand the relationships between the latent or space process $\{\epsilon_t\}$ and the zero-inflated count process $\{Y_t\}$ and their corresponding autocorrelation functions. In Subsection 3.5.2, we provide more comprehensive simulation study to evaluate the estimation method proposed in this chapter.

3.5.1 EXAMPLE I

In this simulated example, we generate zero-inflated count process following the ZIP distributions through an $AR(1)$ process, and study the relationship between the two processes.

Based on the proposed model given in (26), the zero-inflated count process is generated as follow.

1. Generate normally distributed process such that:

   (a) $\epsilon_t \sim N(0,1)$ for $t = 1, \ldots, n$.

   (b) $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, R(\varphi))$ where $R(\varphi)$ is an $AR(1)$ correlation matrix

2. Compute $u_t = \Phi(\epsilon_t)$, for $t = 1, \ldots, n$, where Φ is the cdf of standard normal distribution.

3. Compute $y_t = F_t^{-1}(u_t)$, i.e. choose the smallest value of $y_t \in 0, 1, 2, \ldots$ for which the cdf of y_t is greater than or equal to u_t, i.e. $F_t(y_t) \geq u_t$, where F_t is the cdf of ZIP distribution.

Figure 5 shows the densities of the zero-inflated counts $\{Y_t\}$ following the ZIP marginals with $\lambda = 4.3$ and $\omega = 0.25$ generated through the normally distributed errors $\{\epsilon_t\}$ which follows an $AR(1)$ process with $\varphi = 0.35.$ As discussed in Section 3.2, the construction of the model given in (26), ensures that the process $\{Y_t\}$ follows the desired ZIP marginal by the integral transformation theorem, which can be seen in the left graph of Figure 5, and the errors are normally distributed as seen in the right graph of Figure 5.

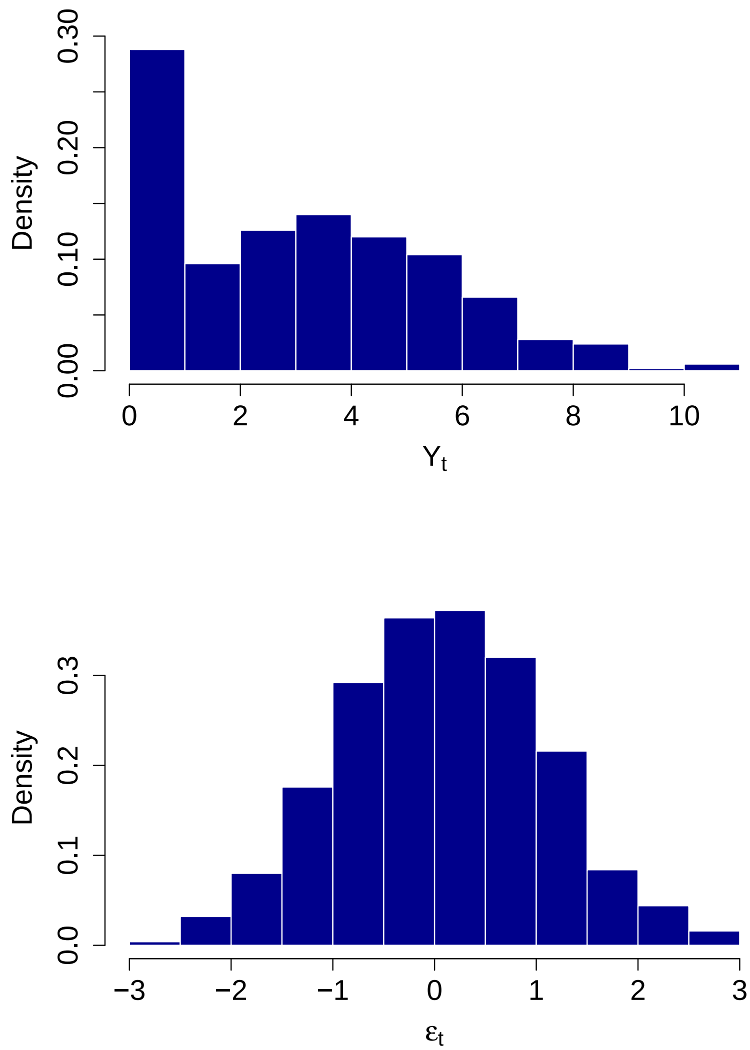


Figure 5: Histograms of the zero-inflated counts $\{Y_t\}$ following the ZIP marginals with $\lambda = 4.3$ and $\omega = 0.25$ (top), and the underline process $\{\epsilon_t\}$ following $AR(1)$ with $\varphi = 0.35$ (bottom)

In addition, we plot the zero-inflated counts $\{Y_t\}$ against $\{\epsilon_t\}$ and $\{u_t\}$ where $u_t = \Phi(\epsilon_t)$ for $t = 1, \ldots, n$ in Figure 6. Due the zero-inflation in the counts, the ranges of ϵ_t and p_t at $Y_t = 0$ are wider. In fact, $Y_t = 0$, whenever $u_t \in (0, P(Y_t = 0|\lambda = 4.3, \omega = 0.25)]$ or equivalently $\epsilon_t \in (0, \Phi^{-1}(P(Y_t = 0|\lambda = 4.3, \omega = 0.25))]$. At $Y_t > 0$, the ranges of ϵ_t and u_t is controlled by the intensity parameter $\lambda = 4.3$, which explains why they are relatively wider around λ than further from it. This explains how the zero-inflated process is generated by the discretization of the latent Gaussian ARMA process.

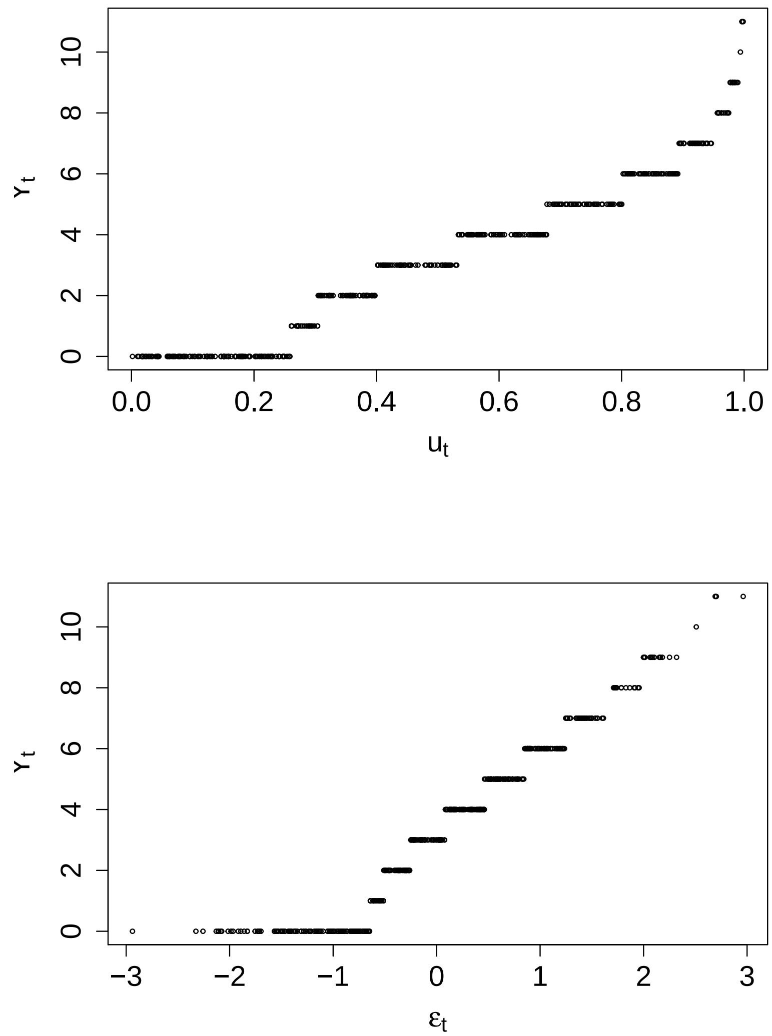
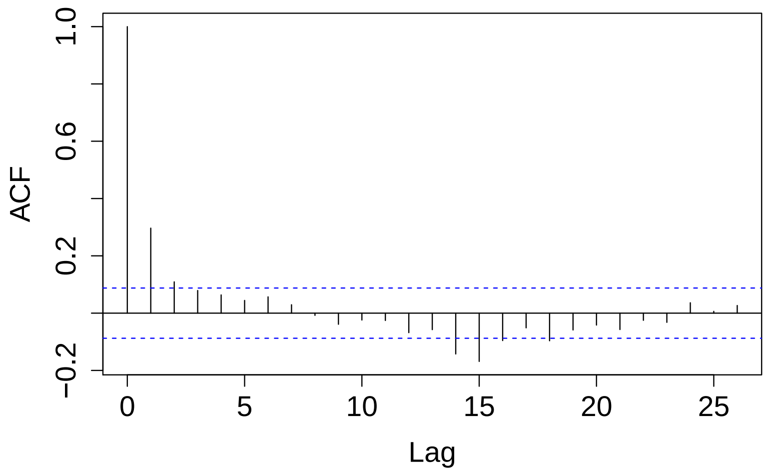


Figure 6: scatter plots of $\{Y_t\}$ against $\{u_t\}$ where $p_t = \Phi(\epsilon_t)$ (top), and $\{Y_t\}$ against $\{\epsilon_t\}$ (bottom).

Figure 7 shows the sample autocorrelation functions (ACFs) of the two processes $\{Y_t\}$ and $\{\epsilon_t\}$. There is clear similarity between the two sample ACFs in which $|\rho_Y(h)|$ is slightly less than $|\rho_\epsilon(h)|$. This agrees with the finding discussed in Section 3.2.1. In fact, using the link function defined in (31), and given that $\rho_Y(1) = \varphi = 0.35$, the first order autocorrelation corresponding to the zero-inflated counts $\rho_Y(1)$ equals to 0.33.



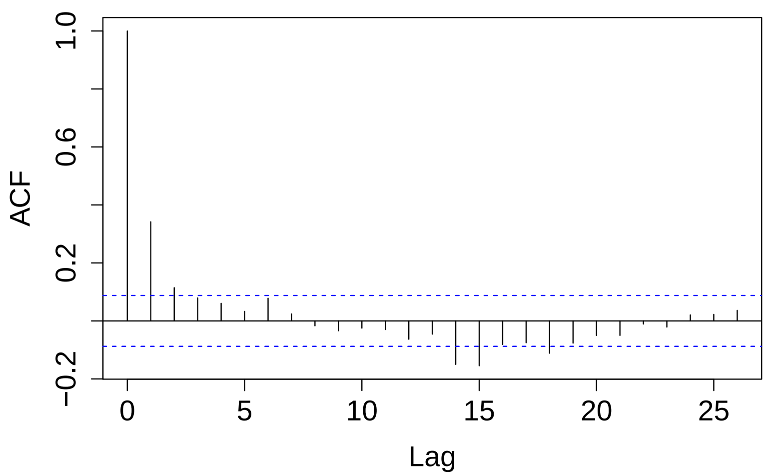


Figure 7: Autocorrelation function (ACF) of the zero-inflated count process $\{Y_t\}$ (top) and ACF of the error process $\{\epsilon_t\}$ (bottom).

3.5.2 EXAMPLE II

In this example, we run the same simulation algorithm as in Example I but with different sample sizes from the ZIP, ZINB, and ZICMP marginals and under the $AR(1)$ and $MA(1)$ dependence structures to evaluate performance of the estimation method. First, we consider the $AR(1)$ dependence structure. Only one covariate X_t is considered and chosen to be the same for the intensity parameter across all distributional assumptions. The covariate is

given by

$$x_t = 0.6x_{t-1} - 0.4x_{t-2} + \zeta_t, \quad \zeta_t \overset{\text{i.i.d.}}{\sim} N(0,1),$$

that is, the log linear model is given by

$$\log(\lambda_t) = \beta_0 + \beta_1 x_t,$$

whereas the zero-inflated and dispersion parameters are chosen to be constant across time when we simulate from all the distributions and ZINB and ZICMP distributions for the dispersion parameter, that is, $\omega_t = \omega$ and $\kappa_t = \kappa$ for $t = 1, \ldots, n$. We consider the following models under stationary $AR(1)$ errors with the parameter $\varphi = 0.5$:

- ZIP with $\boldsymbol{\beta} = (4.3, 0.3)'$ and $\omega = 0.25$;
- ZINB with $\boldsymbol{\beta} = (4.3, 0.3)'$, $\omega = 0.2$ and $\kappa = 0.5$;
- ZICMP with $\boldsymbol{\beta} = (5, 0.3)'$, $\omega = 0.2$ and $\kappa = 0.5$.

We generate 500 simulated datasets for each of the above models with the sample sizes, $n = 100, 200$ and 500. The evaluation criterion is chosen to be the mean absolute deviation error (MADE), which is given by

$$\frac{1}{m}\sum_{i=1}^{m} |\widehat{\vartheta}_i - \vartheta|,$$

where m is the number of replications, i.e. $m = 500$.

The parameter estimates were obtained using the R package "gcmr" (Masarotto and Varin, 2017) after constructing our own codes for the marginal models of the ZIP, ZINB, and ZICMP distributions. A summary of the simulation results are shown in Table 1, which represents the count time series ZIP, ZINB, and ZICMP models with stationary $AR(1)$ errors. The results indicate that the proposed estimation method produces reasonable estimates and relatively small MADEs. In addition, as the sample size increases the parameter estimates seem to converge to the true parameter values.The box plots displayed in Figures 8, 9, and 10 show how the performance enhanced when the sample size increased. Although we did not discuss the asymptotic properties of the parameter estimates obtained from the simulated likelihood function, several authors (see for examples, Gourieroux and Monfort, 1990 and Lee, 1999) proved that such methods provide normally distributed estimates. To

assess the approximate normality of the estimates, Q-Q plots of the ML estimates for the 500 ZIP, ZINB and ZICMP replicates of length $n = 500$ are shown in Figures 11, 12, and 13.

Table 1: Mean of estimates, MADEs (within parentheses) for zero-inflated models with AR(1) dependence structure.

Model	n	β_0	β_1	ω	κ	φ
ZIP	100	4.3012(0.0159)	0.2998(0.0071)	0.2448(0.0524)		0.4847(0.0716)
	200	4.3008(0.0111)	0.3000(0.0056)	0.2518(0.0545)		0.4945(0.0479)
	500	4.3004(0.0074)	0.3000(0.0038)	0.2505(0.0446)		0.4985(0.0293)
ZINB	100	4.2597(0.2068)	0.3002(0.0872)	0.1946(0.0581)	0.5470(0.1060)	0.4702(0.0809)
	200	4.2797(0.1382)	0.3015(0.0667)	0.1980(0.0408)	0.5245(0.0714)	0.4864(0.0533)
	500	4.2904(0.0912)	0.2987(0.0457)	0.1993(0.0238)	0.5121(0.0439)	0.4936(0.0329)
ZICMP	100	5.3744(0.7838)	0.3214(0.0482)	0.1956(0.0690)	0.5568(0.1477)	0.4726(0.0783)
	200	5.2059(0.5138)	0.3124(0.0341)	0.1986(0.0581)	0.5329(0.0329)	0.4868(0.0526)
	500	5.1010(0.3234)	0.3059(0.0216)	0.1994(0.0513)	0.5167(0.0167)	0.4949(0.0319)

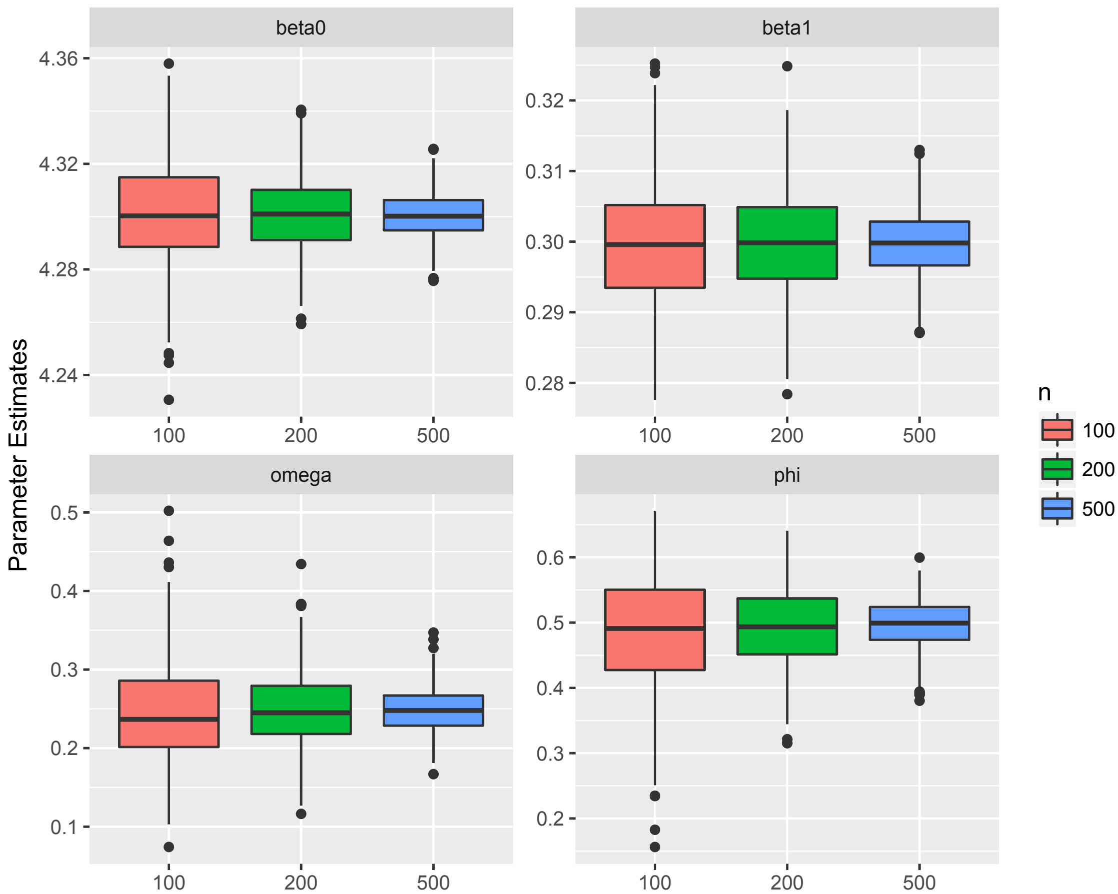


Figure 8: ML estimates for the 500 ZIP-AR(1) models of length $n = 100, 200,$ and 500

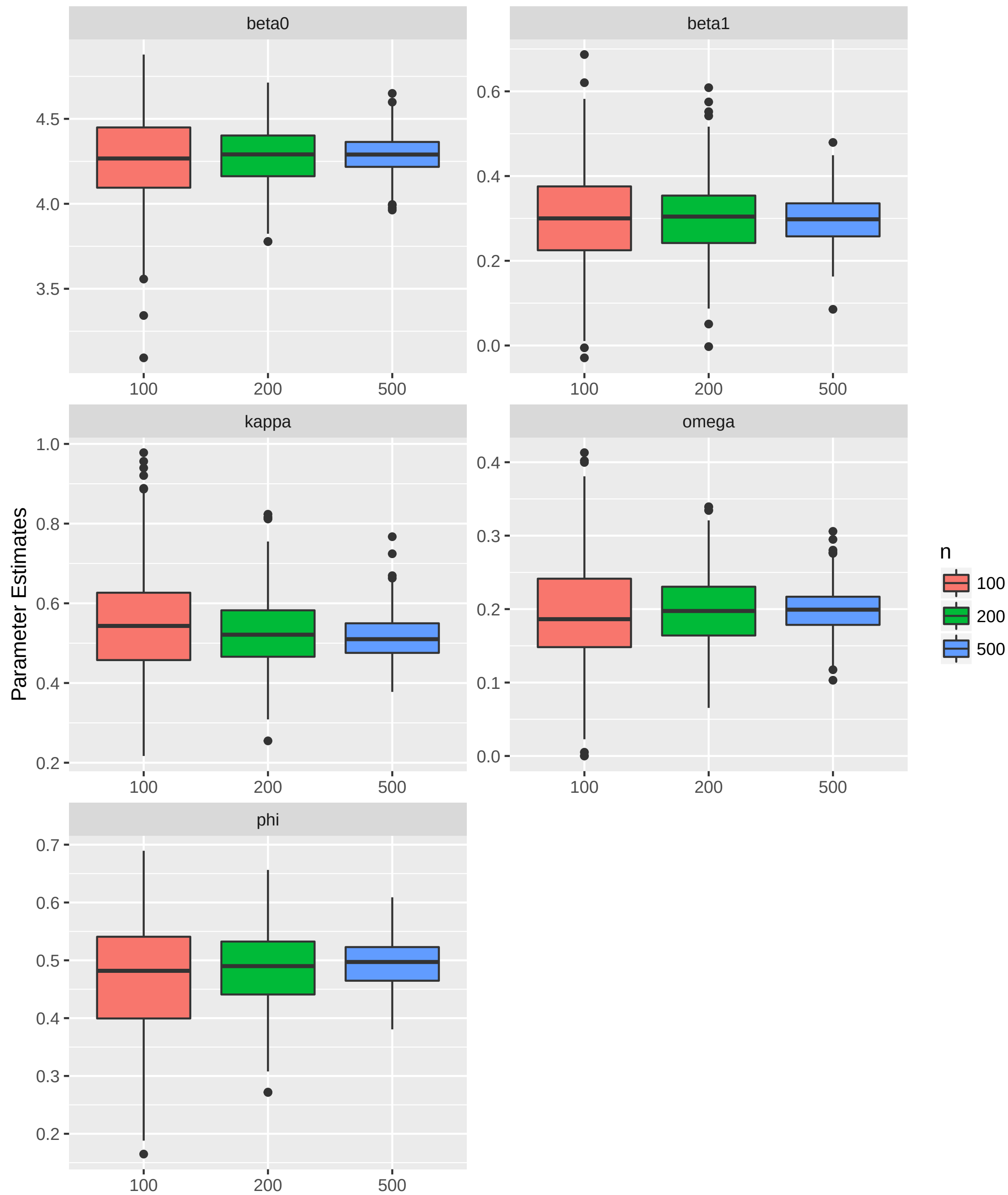


Figure 9: ML estimates for the 500 ZINB-AR(1) models of length $n = 100, 200,$ and 500

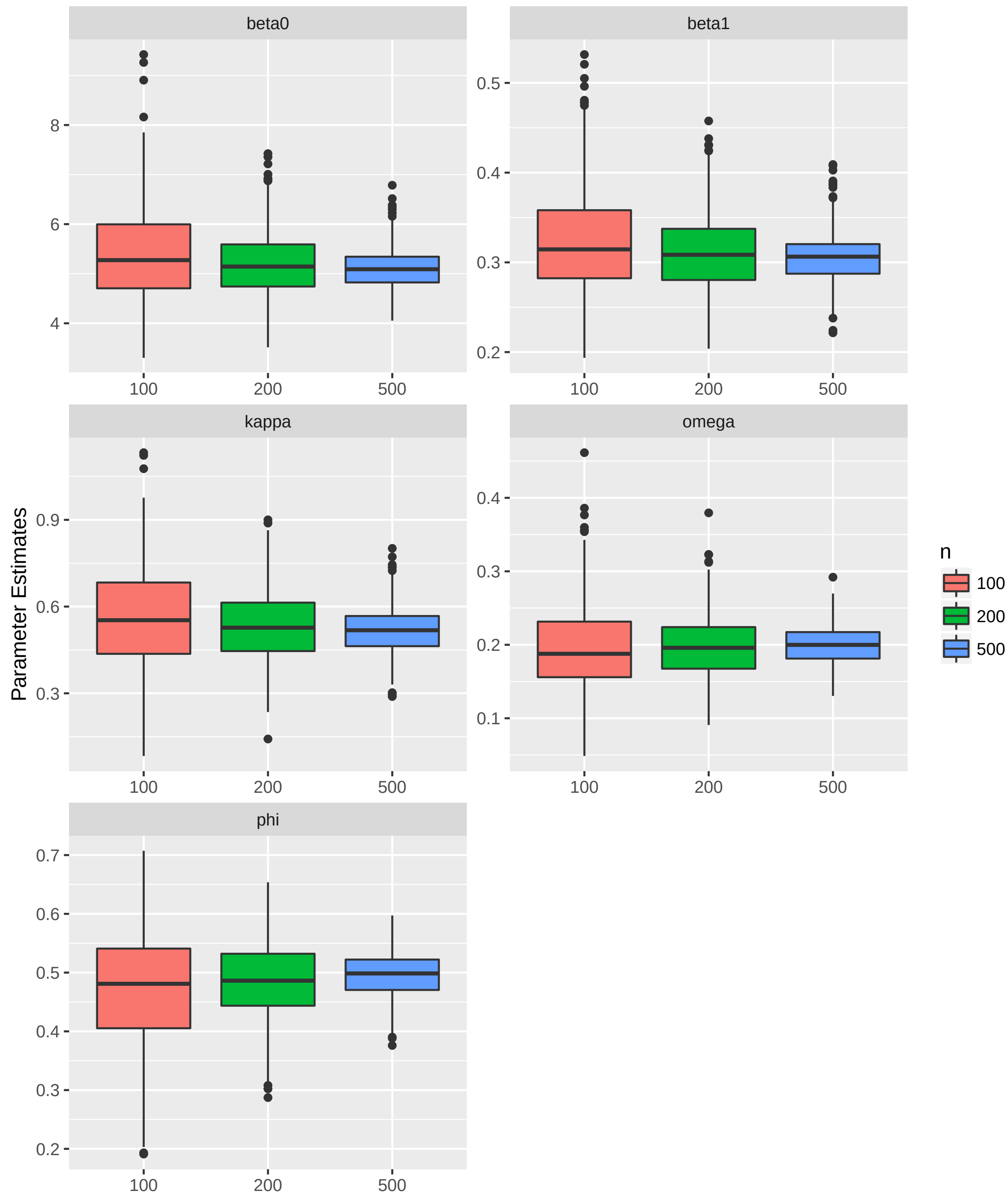


Figure 10: ML estimates for the 500 ZICMP-AR(1) models of length $n = 100, 200,$ and 500

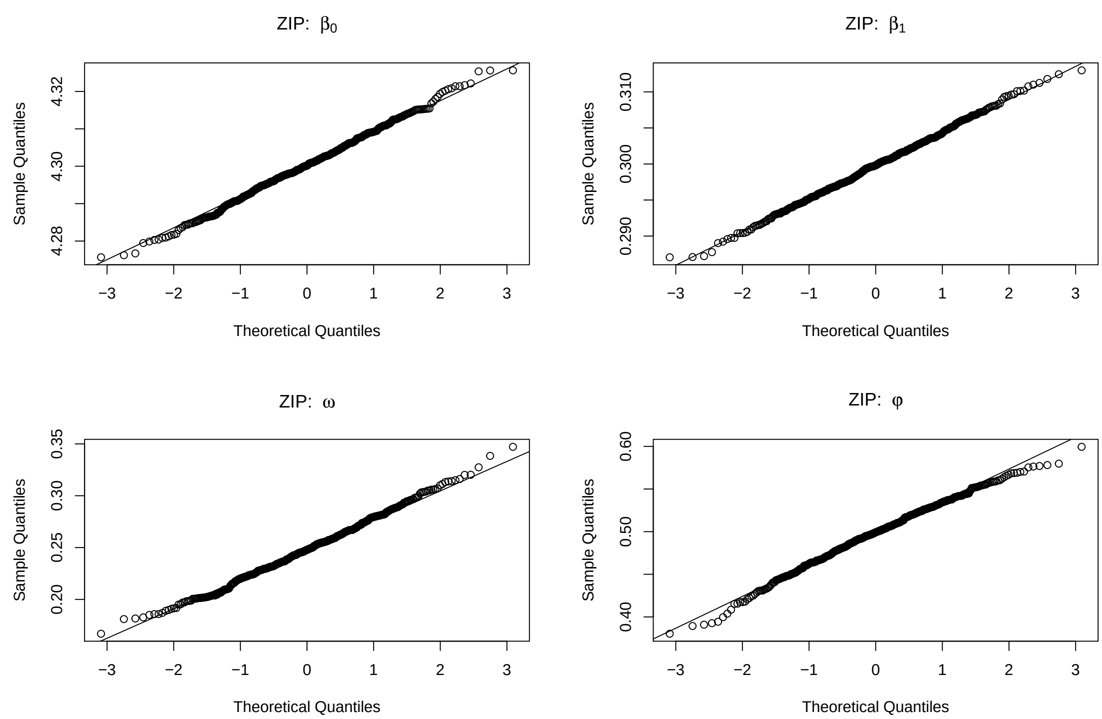


Figure 11: Q-Q plots of the ML estimates for the 500 ZIP-AR(1) process of length $n = 500$.

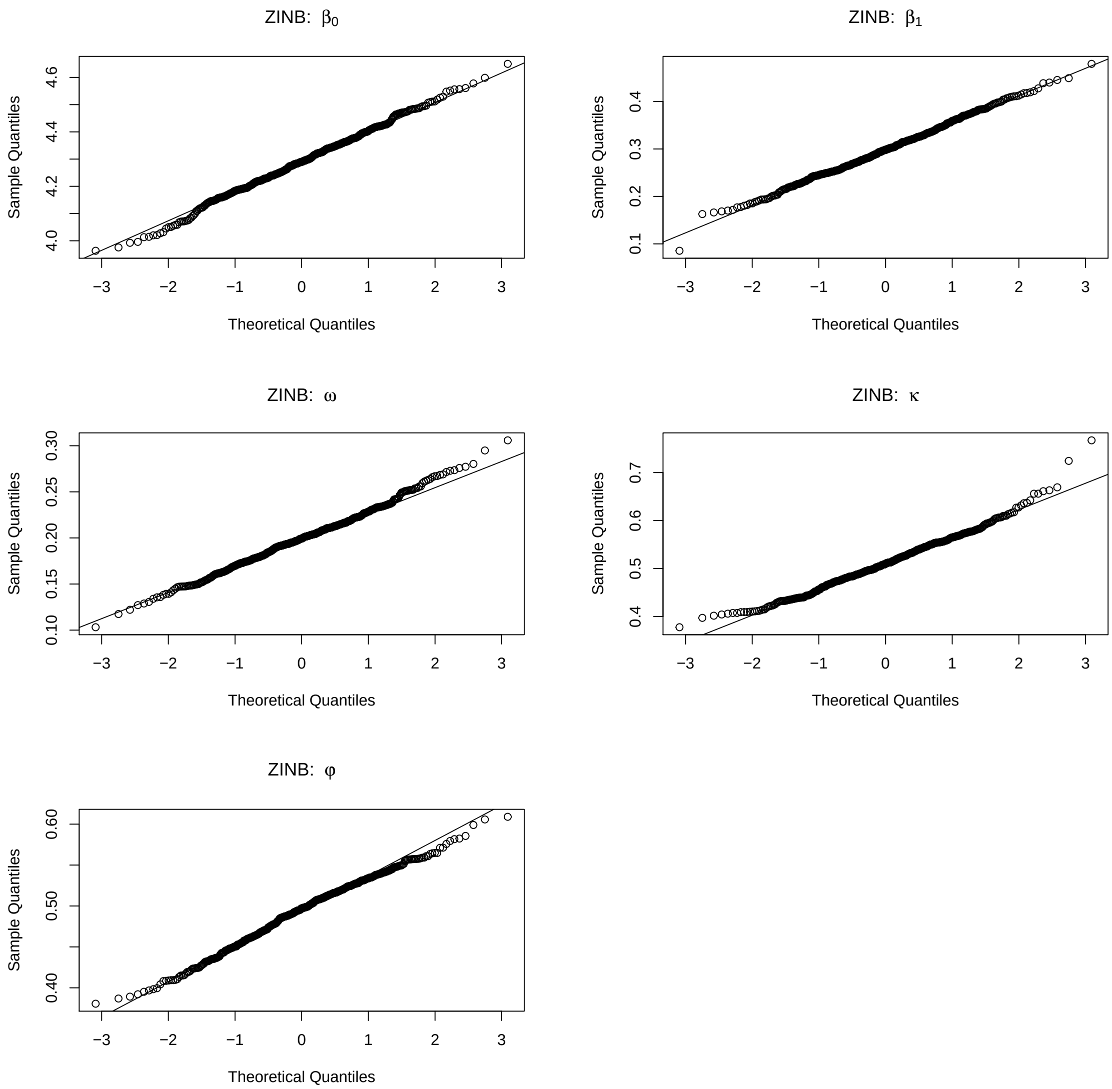


Figure 12: Q-Q plots of the ML estimates for the 500 ZINB-AR(1) process of length $n = 500$.

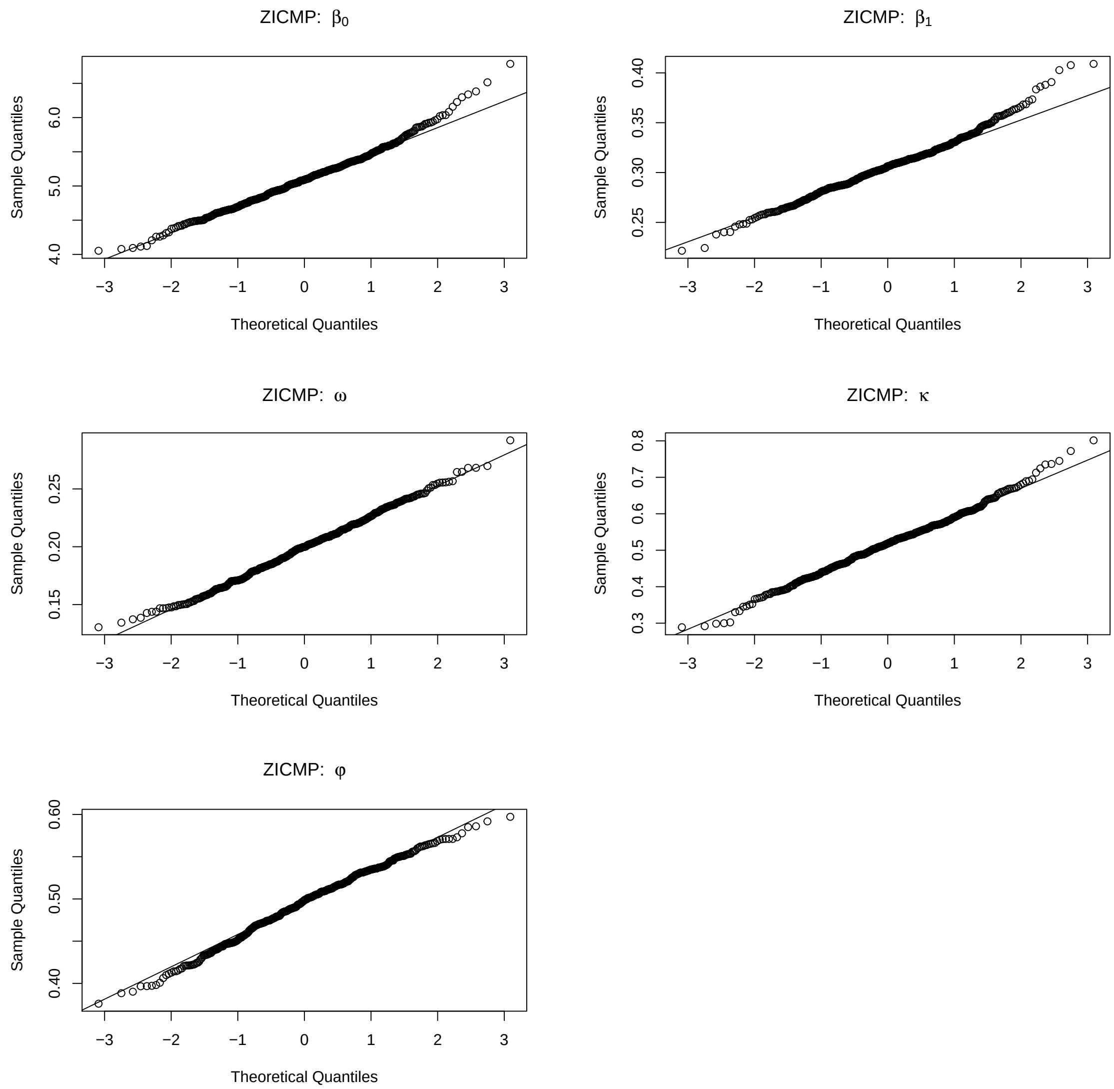


Figure 13: Q-Q plots of the ML estimates for the 500 ZICMP-AR(1) process of length $n = 500$.

Second, we consider the $MA(1)$ dependence structure. No covariates are considered, so the intensity parameter λ, zero-inflated parameter ω, and the dispersion parameter κ (if existed) are constant across time. The dependence parameter of the latent $MA(1)$ process is chosen to be $\delta = 0.5$ across all three marginals. The marginal parameters are then given by

- ZIP with $\lambda = 4.3$ and $\omega = 0.25$;
- ZINB with $\lambda = 4.3$, $\omega = 0.25$ and $\kappa = 0.5$;
- ZICMP with $\lambda = 3'$, $\omega = 0.2$ and $\kappa = 0.25$.

Table 2 shows a summary of the simulation results for the ZIP, ZINB, and ZICMP models with with stationary $MA(1)$ errors. The summary shows that the proposed estimation method performs well with the latent process $\{\epsilon_t\}$ following MA(1) process. The box plots displayed in Figures 14, 15, and 16 show the increase in the dimension enhances the performance of the estimation method. Also, the estimates seem to be asymptotically normally distributed, which can be seen in Figures 17, 18, and 19.

Table 2: Mean of estimates, MADEs (within parentheses) for zero-inflated models with MA(1) dependence structure.

Model	n	λ	ω	κ	δ
ZIP	100	4.3223(0.1903)	0.2533(0.0347)		0.5167(0.0869)
	200	4.3205(0.1290)	0.2521(0.0272)		0.5038(0.0578)
	500	4.3088(0.0933)	0.2514(0.0176)		0.4961(0.0368)
ZINB	100	4.4940(0.8777)	0.2462(0.1281)	0.6227(0.2387)	0.5289(0.1102)
	200	4.3317(0.7072)	0.2293(0.1099)	0.5556(0.1712)	0.4958(0.0661)
	500	4.2904(0.4367)	0.2413(0.0721)	0.5305(0.1041)	0.4955(0.0456)
ZICMP	100	3.2587(0.5566)	0.3421(0.1421)	0.2585(0.0491)	0.5119(0.0842)
	200	3.1520(0.3189)	0.3400(0.1400)	0.2544(0.0375)	0.4992(0.0554)
	500	3.1229(0.2183)	0.3397(0.1397)	0.2549(0.0241)	0.4978(0.0366)

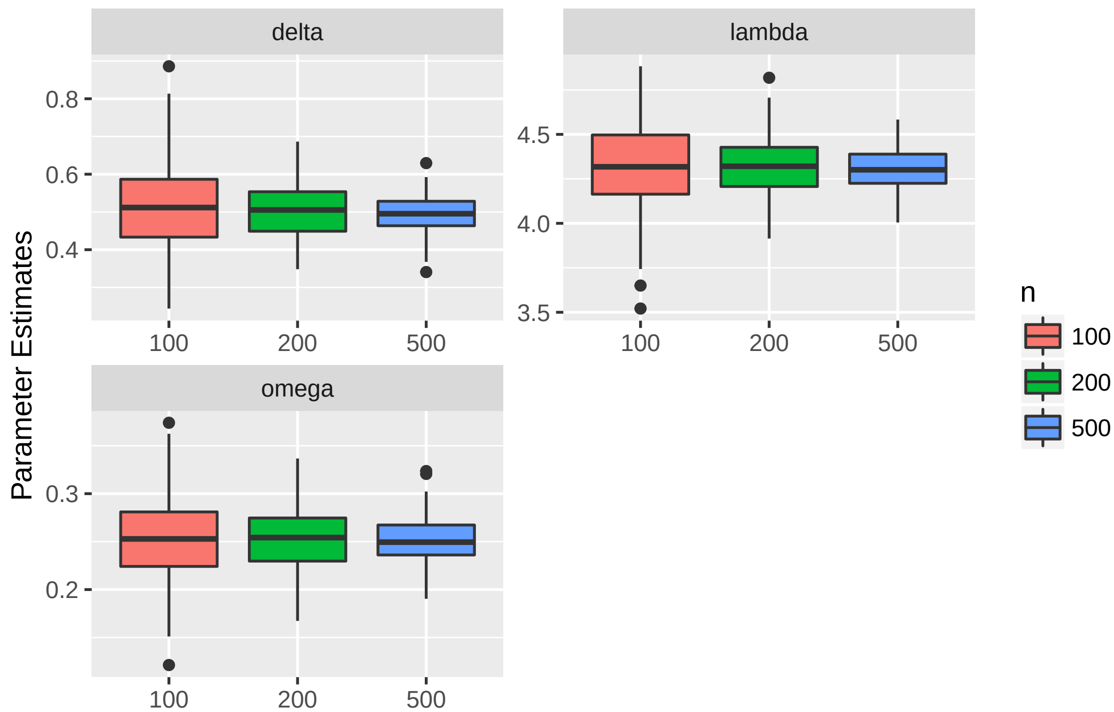


Figure 14: ML estimates for the 500 ZIP-MA(1) models of length $n = 100, 200,$ and 500

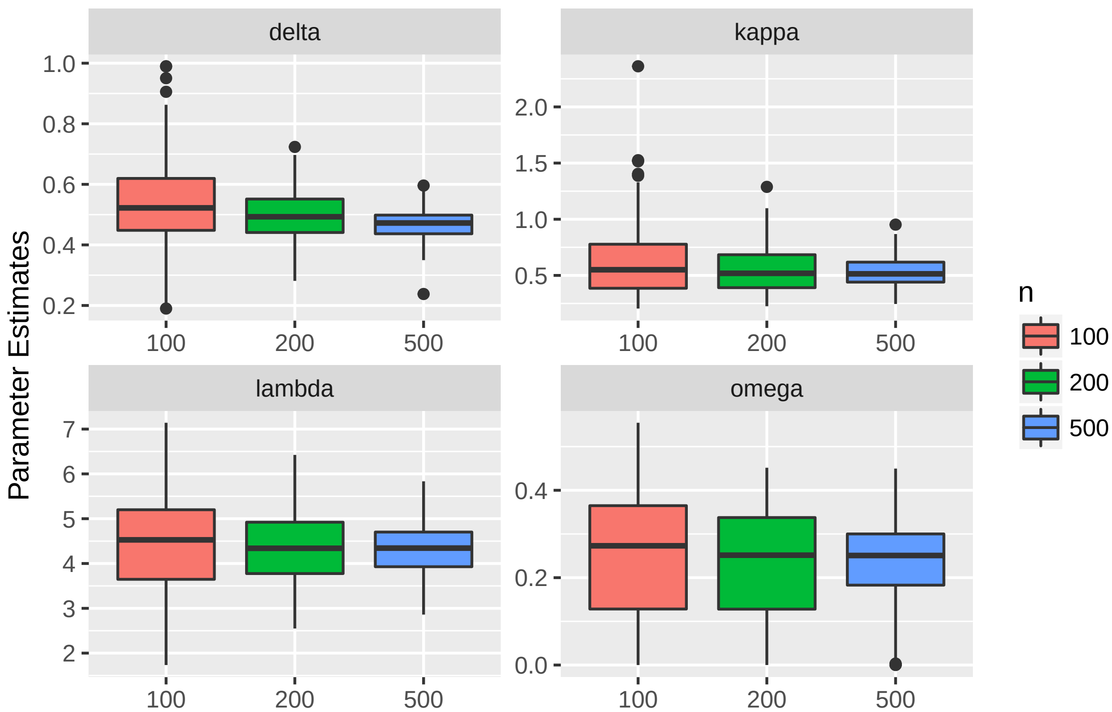


Figure 15: ML estimates for the 500 ZINB-MA(1) models of length $n = 100, 200,$ and 500

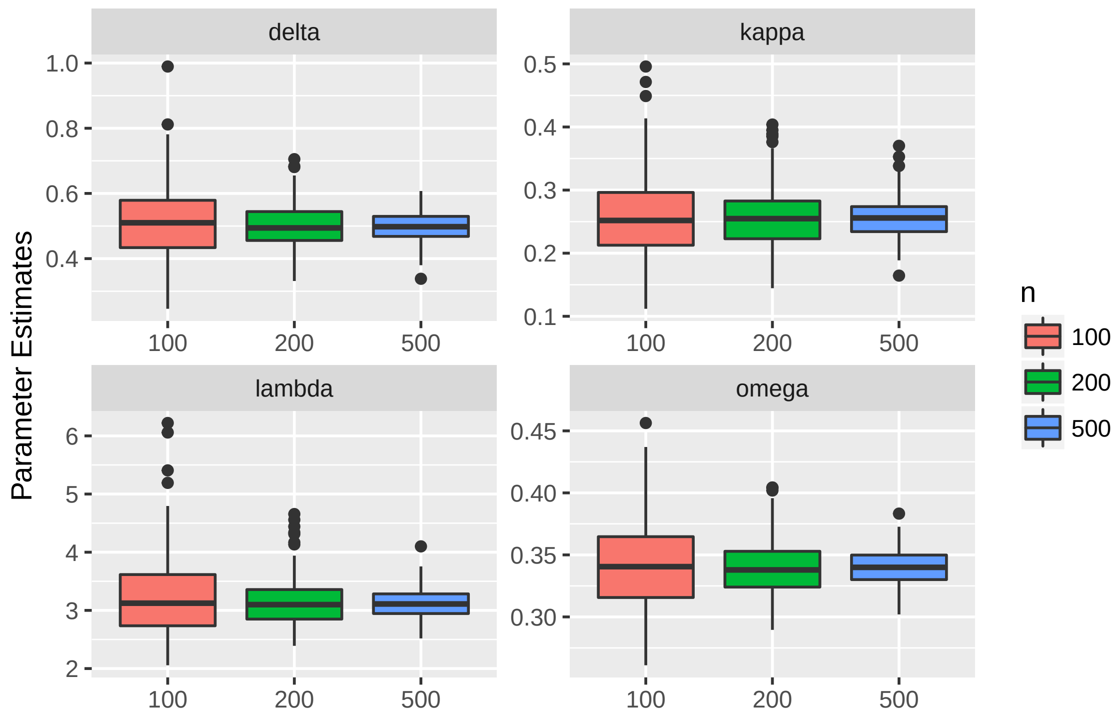


Figure 16: ML estimates for the 500 ZICMP-MA(1) models of length $n = 100, 200,$ and 500

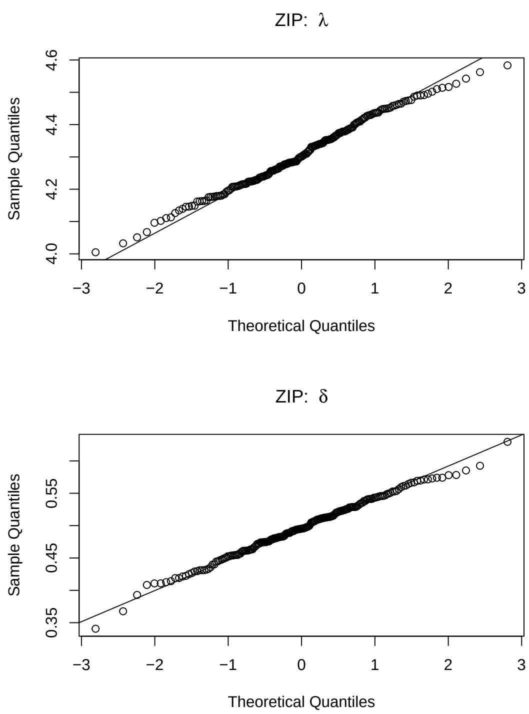


Figure 17: Q-Q plots of the ML estimates for the 500 ZIP-MA(1) process of length $n = 500$.

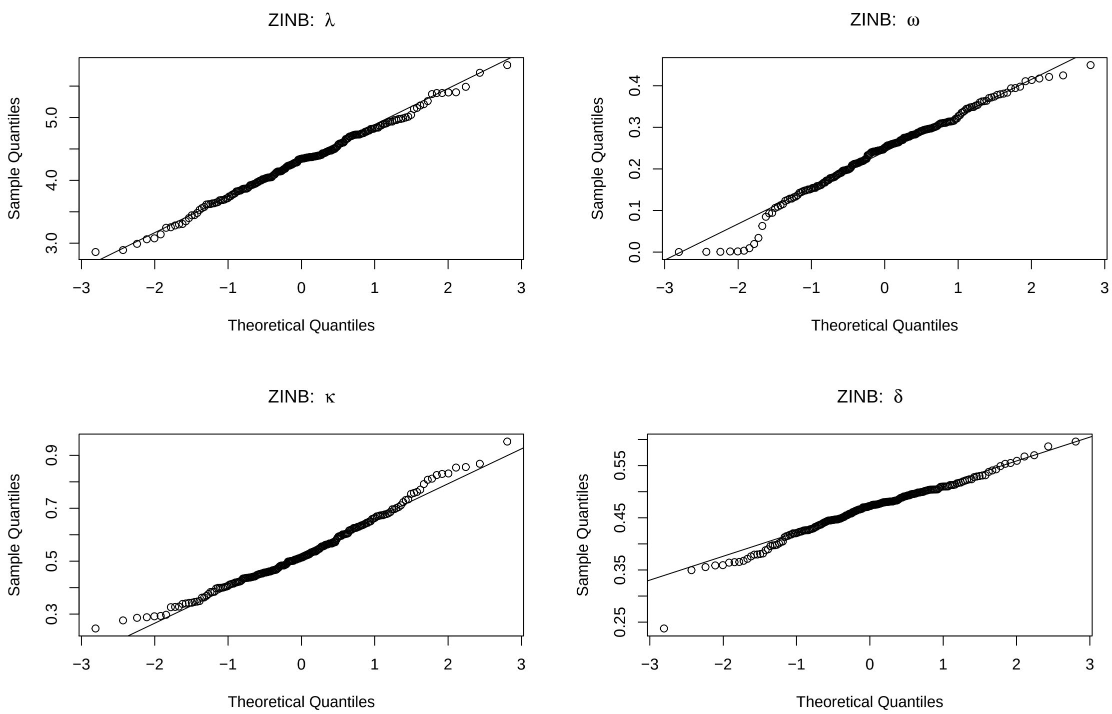


Figure 18: Q-Q plots of the ML estimates for the 500 ZINB-MA(1) process of length $n = 500$.

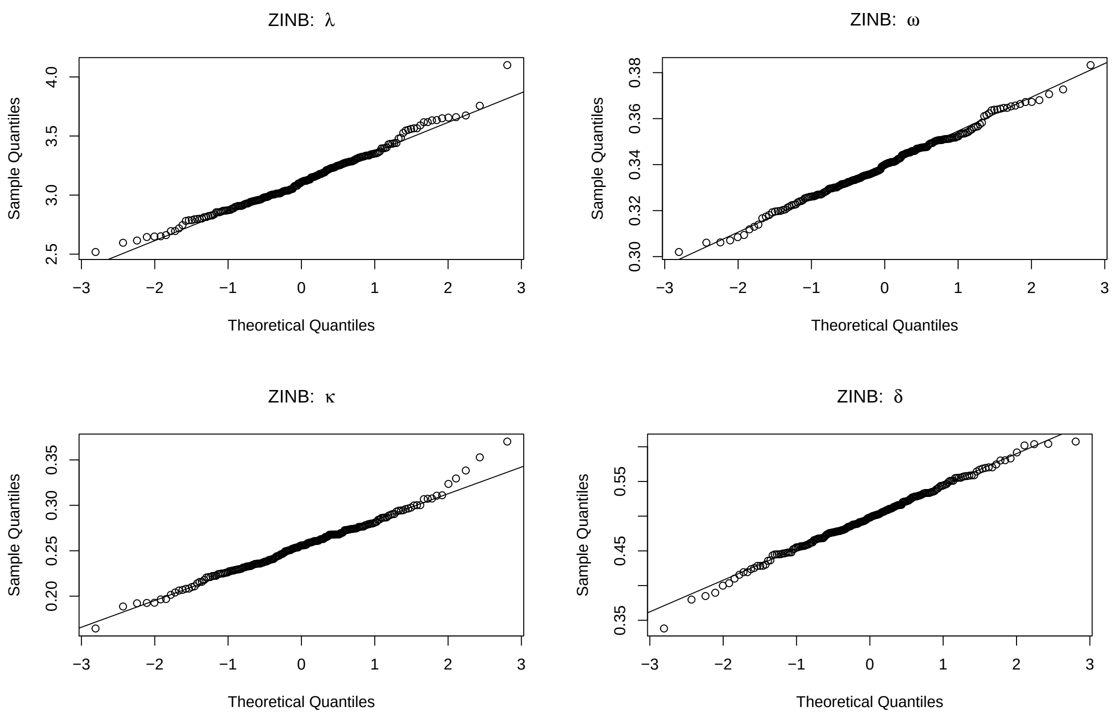


Figure 19: Q-Q plots of the ML estimates for the 500 ZICMP-MA(1) process of length $n = 500$.

3.6 APPLICATIONS

3.6.1 INJURY DATA

In this section, we applied the proposed models using the occupational health data presented in Yau et al. (2004) and Yang et al. (2015). The application concerns the assessment of a participatory ergonomics intervention in reducing the incidence of work place injuries among a group of cleaners in a hospital. The data consists of 96 monthly counts of work-related injuries, starting from July 1988 and ending in October 1995. The participatory ergonomics intervention was commenced on November 1st 1992. That is, 57 observations were pre-intervention and 39 post-intervention. Empirical mean and variance of the time series of counts are 1.4688 and 3.8306, respectively. A bar plot of the distribution of series is displayed in Figure 20, from which we can see that the distribution of the time series of injury counts has more zeros relative to a Poisson distribution with the same empirical mean. Yau et al. (2004) stated that the frequent occurrence of zeros is due to the heterogeneity in risk and the dynamic population of cleaners. The zeros represent about 48% of the sample. The count series and the corresponding sample autocorrelation function of the series are shown in Figure 21. We can see from the plots that there exist frequent occurrence of zeros and low ordered autocorrelation. In addition, the difference between the count series before and after the intervention is intriguing.

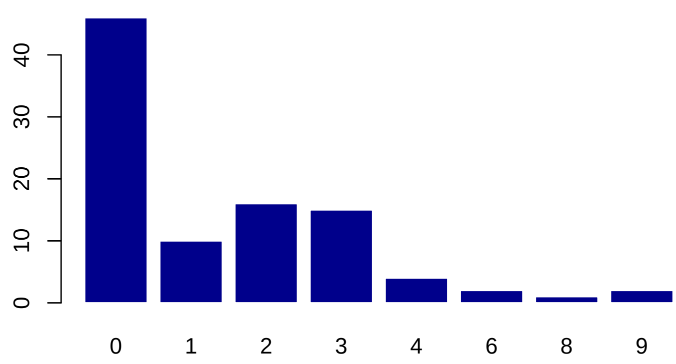


Figure 20: Bar plot of the injury counts series

Hence, we fit several models to investigate whether the participatory ergonomics intervention reduces the injury counts or not. In addition, we study the serial dependence of the counts. The models take the form given in (26) with the following log-linear function for the intensity parameter

$$\log(\lambda_t) = \beta_0 + \beta_1 x_t, \quad t = 1, \ldots, 96,$$

where x_t is a binary variable that takes the value zero if $t < 57$, the intervention time, and one otherwise. The zero inflation parameter, ω, is assumed to be constant across time. The same is assumed on the dispersion parameter, κ, when we take the ZINB and ZICMP distributions. Thus, the main model is given by

$$Y_t = F_t^{-1}\{\Phi(\epsilon_t)|x_t; \boldsymbol{\theta}\}, \quad t = 1, \ldots, 96$$

where $\boldsymbol{\theta} = (\beta_0, \beta_1, \gamma, \alpha)'$ with $\gamma = \text{logit}(\omega)$ and $\alpha = \log(\kappa)$. The latent random process, the errors, are generally given by the $ARMA(p, q)$ process. However, here and after fitting multiple models, we consider only those with the errors following $AR(1)$ process, which correspond to the smallest Akaike information criterion (AIC) (Akaike, 1974).

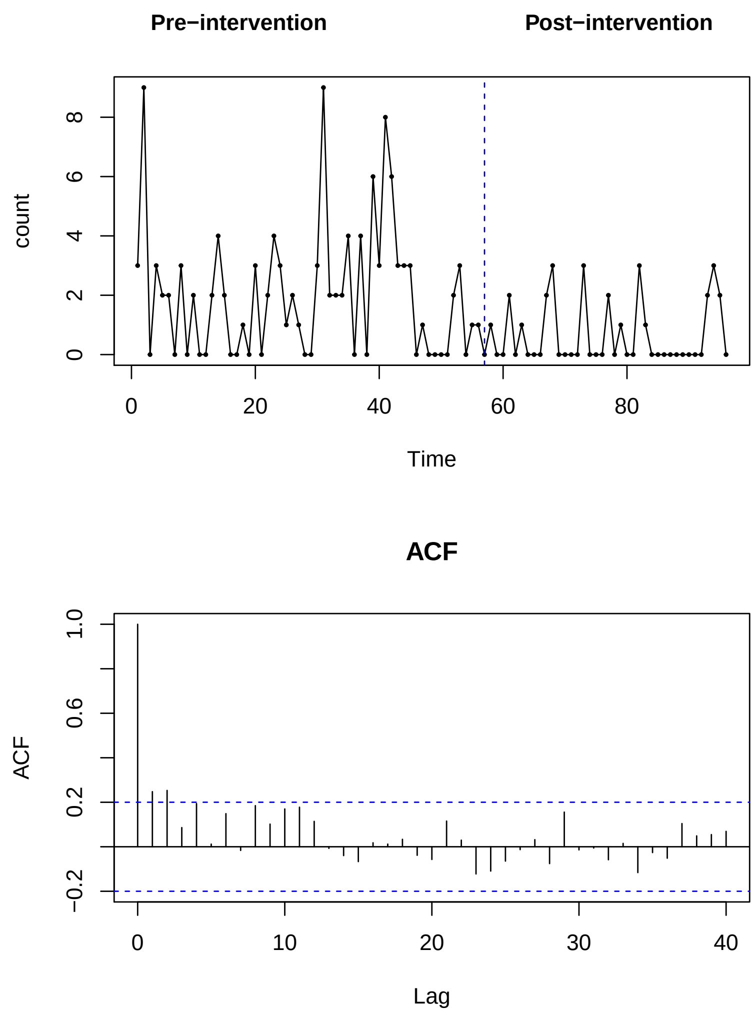


Figure 21: Injury counts series: the time series plot and the sample autocorrelation.

Table 3 shows the three copula-based zero-inflated models we proposed in this chapter along with the copula-based Poisson and NB models introduced in Masarotto and Varin (2012) all with the $AR(1)$ correlation structure. The Poisson model seems to perform moderately less than the other models because it fails to account for the overdispersion in the counts caused by the zero inflation. On the other hand, adding more probability to the event zero improves the performance of the fitted model since it accounts for the overdispersion and the frequent occurrence of zeros. This is why the ZIP, ZINB and ZICMP models perform better in term of AIC than the Poisson and NB models. All five models suggest that the work-related injuries significantly decreased after implementing the participatory ergonomics intervention since the value of $\widehat{\beta}_1$ is always less than zero.

However, the degree of significance of $\widehat{\beta}_1$ is less prominent after accounting for the zero inflation.

Table 3: Parameter estimates (standard errors) for the copula-based models fit to the injury count series.

Parameter	ZIP	ZINB	ZICMP	Poisson	NB
β_0	1.0794 (0.1044)	1.0282 (0.1398)	0.3611 (0.3751)	0.7148 (0.1019)	0.6945 (0.1730)
β_1	-0.8605 (0.3065)	-0.9410 (0.3187)	-0.6981 (0.2429)	-1.0989 (0.2326)	-1.0837 (0.3219)
γ	-0.5180 (0.3036)	-0.7492 (0.3951)	-0.9074 (0.4513)		
α		0.7185 (0.6340)	0.4785 (0.5149)		0.9557 (0.3135)
φ	0.1201 (0.1117)	0.1186 (0.1222)	0.1227 (0.0889)	0.1012 (0.0695)	0.1000 (0.1183)
AIC	310.02	308.42	308.64	345.66	313.27

To examine the model assumptions, we first test the assumption that the latent process $\{\epsilon_t\}$ has a unit variance. Since the process is serially dependent, standard variance tests are not appropriate because they are sensitive to dependency. An appropriate method to test whether $\sigma_\epsilon = 1$ or not would be in using the stationary block bootstrap (SBB) to draw confidence interval (CI) of the variance (Politis and Romano, 1994 and Lahiri, 2013). Table 4 shows the estimated value of σ_ϵ for zero-inflated marginals and the ordinary Poisson and NB marginals. After estimating the latent process using (49), we draw SBB 95% CIs for σ_ϵ under each marginals. The results show that 1 is included in all the CIs except for the Poisson marginal, which suggests that choosing the Poisson leads to violation of the assumption.

Table 4: SBB 95% CI of ϵ variance for the injury count series.

Model	$\widehat{\sigma}_\epsilon$	SBB 95% CI
Poisson	1.6589	(1.2040, 2.1560)
NB	1.1314	(0.9120, 1.3810)
ZIP	0.9306	(0.6741, 1.2104)
ZINB	0.8394	(0.6636, 1.0336)
ZICMP	1.1544	(0.9310, 1.4050)

Figure 22 features the randomized quantile residuals in normal probability and autocorrelation plots of the copula-based ZIP, ZINB and ZICMP models. The normal probability plots suggest the randomized quantile residuals of these three models follow the normal

distribution, and the autocorrelation plots indicate the absence of the serial dependence in the residuals. These findings suggest that the proposed models in this chapter fit the data adequately. Models with more complicated correlation structures such as $AR(2)$ and $ARMA(1,1)$ were also considered and fitted to the data. No significant improvements were found and thus we recommend using $AR(1)$, which is the correlation structure suggested in both Yau et al. (2004) and Yang et al. (2015).

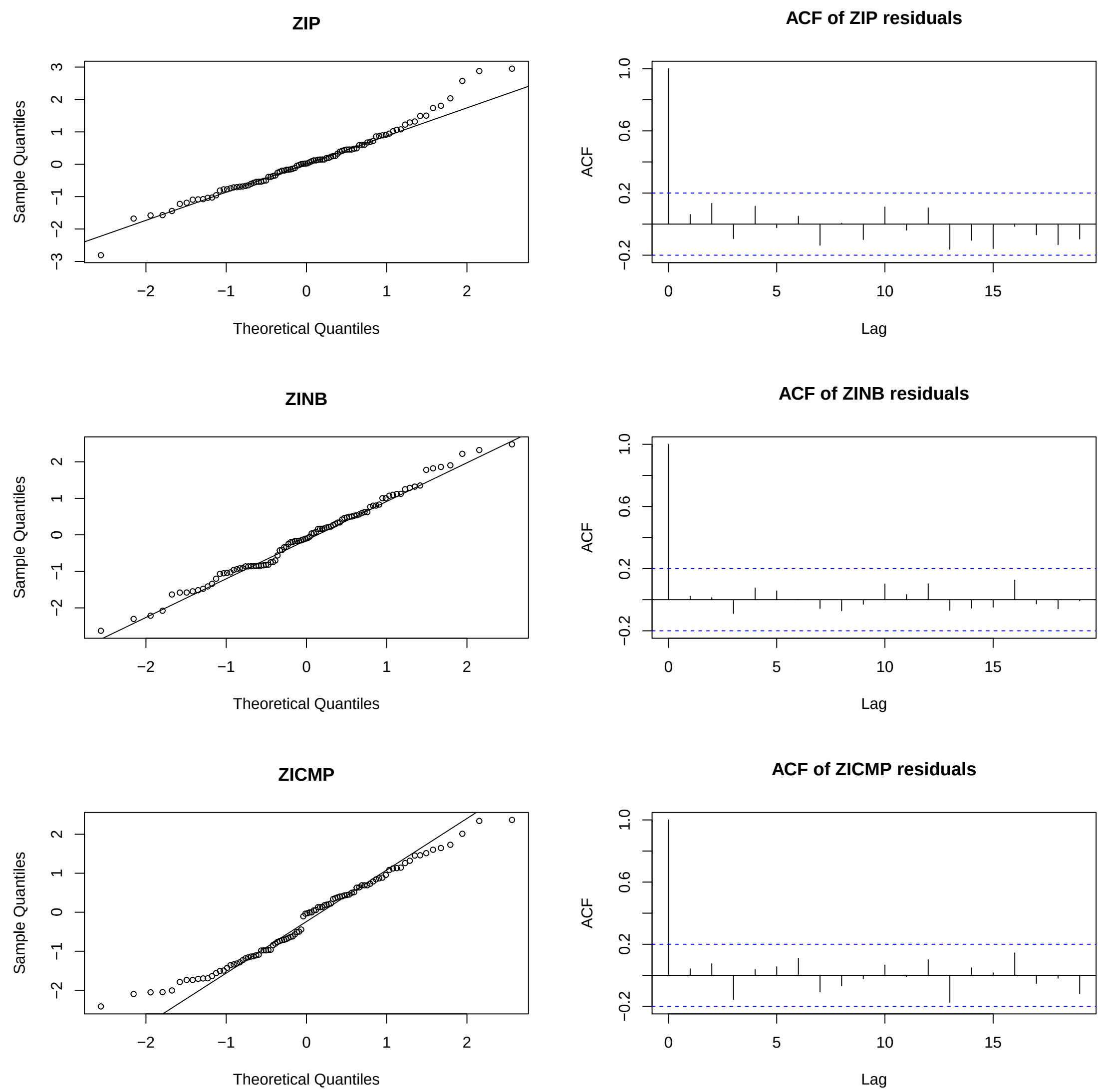


Figure 22: Injury counts series: q-q plots (left) and autocorrelation plots (right) for sets of randomized residuals of the ZIP, ZINB and ZICMP models.

3.6.2 SANDSTORMS DATA

The data set used in this example consists of the monthly count of strong sandstorms recorded by the AQI airport station in Eastern Province, Saudi Arabia. The station happens to be is located in one of the major dust producing regions in the world (Idso, 1976). Sandstorm is a weather event that results from strong wind releasing dust from the ground and transfers it long distances (Goudie and Middleton, 2006). Sandstorms can cause many environmental and human-related hazards. For examples, sandstorms impact the air quality, disturb daily activities, and transportations. Hence, studying and accurately analyzing the behavior of these phenomena is important to successfully forecast such events.

The monthly counts studied here are characterized as strong sandstorms by the AQI airport station. Tao et al. (2002) stated that a strong sandstorm reduces the level of visibility to less than 500 m and with average wind speed of 17.2 to 24.4 m/s. The counts of these events contain zero inflation. Several works have been applied on handling rare events such as strong sandstorms (see for examples Tan et al., 2014 and Ho and Bhaduri, 2015). Here we apply the proposed zero-inflated count time series regression models using Gaussian copula.

The data set consists of 348 monthly counts of strong sandstorms, starting from January 1978 to December 2013. The main objective was to apply the proposed models and investigate if there were any significant seasonal and trend components. Additionally, we investigated if there were any other predictors that affected the frequency of sandstorms such as the monthly counts of dust haze events, maximum wind speed, temperature, and relative humidity.

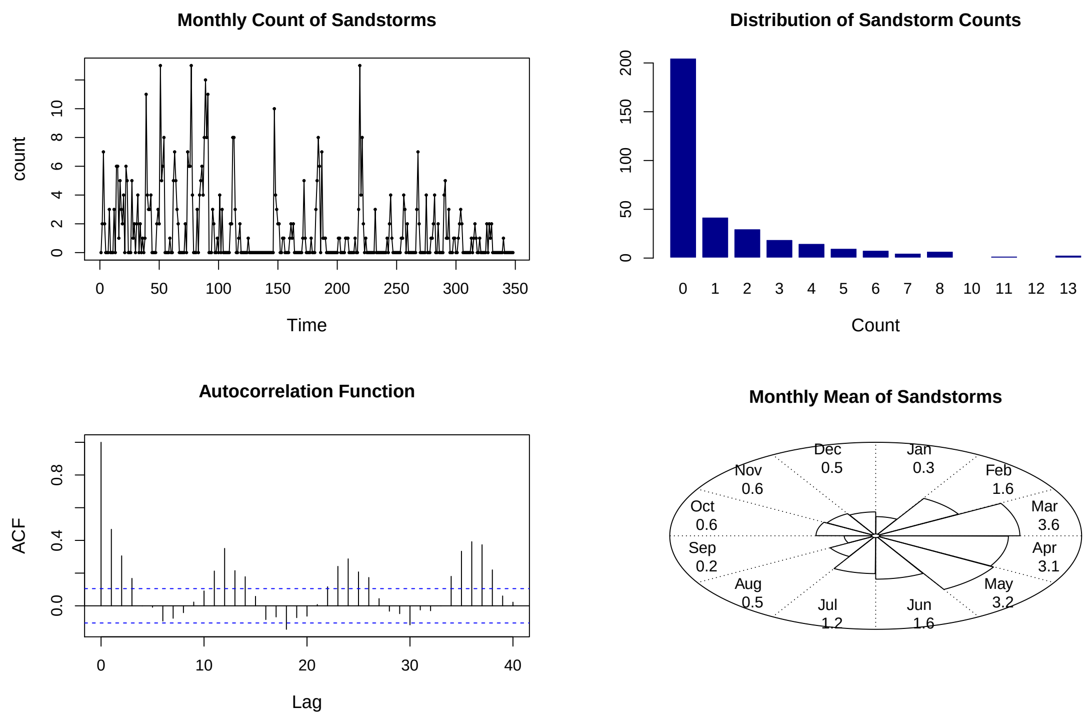


Figure 23: Time series plot of monthly count of sandstorms, the autocorrelation function, bar-plot of distribution of sandstorm counts, and circular plot of the monthly mean count of sandstorms.

Figure 23 shows the sandstorms series plot, the autocorrelation function, bar-plot of the distribution of sandstorm counts, and circular plot of the monthly mean count of sandstorms. From the time series plot and the bar-plot, we could see that the distribution of the sandstorm counts had more zeros relative to a Poisson distribution with the same empirical mean. These zeros represented about 59% of the sample. Decreasing trend could also be observed from the time series plot. Additionally, seasonality was also captured from the autocorrelation function and circular plot. In fact, from the circular plot, we concluded that most sandstorms occurred during spring time, i.e. March, April, and May months. Thus, trend and seasonal covariates were added to the models.

Hence, we fit several models to investigate the trend and seasonality effects along with the other covariates mentioned above. After performing model selection based on AIC, we ended up with the following models taking the form of (26), with the log-linear function of

the intensity parameter given by

$$\log(\lambda_t) = \beta_0 + \beta_1 \left(t \times 10^{-3}\right) + \beta_2 x_{1t} + \beta_3 x_{2t} + \beta_4 x_{3t},$$

and the logit function for the zero-inflation parameter given by

$$\text{logit}(\omega_t) = \gamma_0 + \gamma_1 z_{1t} + \gamma_2 z_{2t} + \gamma_3 z_{3t},$$

for $t = 1, \ldots, n$, where $x_{1t} = z_{1t} = \cos\left(\frac{2\pi t}{12}\right)$, $x_{2t} = z_{2t} = \sin\left(\frac{2\pi t}{12}\right)$, and $x_{3t} = z_{3t}$ is the monthly count of dust haze events. The log-function of the dispersion parameter (if existed) is given by $\log(\kappa) = \alpha$, i.e. it was chosen to be constant across time. Thus, the main model was given by

$$Y_t = F_t^{-1}\{\Phi(\epsilon_t)|\mathbf{X}_t; \boldsymbol{\theta}\}, \quad \text{for } t = 1, \ldots, 348,$$

where $\boldsymbol{\theta} = (\beta_0, \ldots, \beta_4, \gamma_0, \ldots, \gamma_3, \alpha)'$ and $\mathbf{X}_t = (\mathbf{x}_t', \mathbf{z}_t', \mathrm{w}_t)'$, in which the intensity covariates were $\mathbf{x}_t = (t \times 10^{-3}, x_{1t}, x_{2t}, x_{3t})'$, the zero-inflation covariates were $\mathbf{z}_t = (z_{1t}, z_{2t}, z_{3t})'$, and no covariates with the dispersion effect, i.e. $\mathrm{w}_t = 1$ for $t = 1, \ldots, n$. The latent random process, the errors, were generally given by the $ARMA(p, q)$ process. However, after fitting multiple models, we considered the dependence structure that followed $AR(1)$ autocorrelation.

Table 5 shows the three copula-based zero-inflated models we proposed in this chapter along with the copula-based Poisson and NB models introduced in Masarotto and Varin (2012), all with the AR(1) correlation structure. The results of all models are comparable. However, the Poisson and NB model seem to perform moderately less than the other models because they fail to account for the overdispersion in the counts caused by the zero inflation and the zero inflation itself. On the other hand, adding more probability to the event zero improves the performance of the fitted model because it addresses the problem of zero inflation and over dispersion. This is why the ZIP, ZINB, and ZICMP models are better fit than the ordinary Poisson and NB distributions in this application.

Table 5: Parameter estimates (standard errors) for the copula-based models fit to the sandstorms count series.

Parameter	ZIP	ZINB	ZICMP	Poisson	NB
β_0	0.9977(0.1175)	0.9709(0.1570)	0.7978(0.1888)	0.2003(0.1147)	0.2996(0.1965)
β_1	-4.1493(0.6065)	-4.7477(0.7976)	-2.4523(0.5772)	-5.1453(0.5517)	-5.8397(0.9643)
β_2	-0.2004(0.0885)	-0.1813(0.1243)	-0.1089(0.0723)	-0.4634(0.0814)	-0.4385(0.1391)
β_3	0.3461(0.0938)	0.4231(0.1239)	0.2093(0.0786)	0.7879(0.0888)	0.7751(0.1352)
β_4	0.0627(0.0088)	0.0645(0.0123)	0.0435(0.0094)	0.0974(0.0085)	0.0950(0.0163)
γ_0	0.7647(0.2622)	0.5656(0.3047)	0.6629(0.2119)		
γ_1	0.6163(0.2460)	0.6648(0.2925)	-1.0047(0.2132)		
γ_2	-0.8931(0.2401)	-0.8363(0.2736)	-0.1496(0.0344)		
γ_3	-0.1489(0.0424)	-0.1659(0.0524)	-0.2466(0.1613)		
α		0.6400(0.2437)	1.1733(0.2230)		0.9195(0.2009)
φ	0.2580(0.0623)	0.2503(0.0724)	0.2870(0.078)	0.1539(0.0419)	0.2488(0.0740)
AIC	910.9	**895.62**	905.6	1017.3	923.06

Furthermore, Table 5 shows that the zero-inflated models are capable of accounting for first order autocorrelations. The autocorrelation coefficients, $\widehat{\varphi}$'s, are similar across models although the zero-inflation models suggest stronger autocorrelation among the observations. For the marginal parameters, $\boldsymbol{\theta}$, the estimates are quite similar between the ZIP and ZINB, and slightly different from the ZICMP. All models suggest significant decreasing trend in the number of strong sandstorms since $\beta_1 < 0$. Seasonality also significant at annual frequencies since $\beta_2, \beta_3, \gamma_1$ and γ_2 are significantly different from zero. Finally, the affect of dust haze is significant since both β_4 and γ_3 are significantly different from zero.

Table 6: SBB 95% CI of ϵ variance for the sandstorm series.

Model	$\widehat{\sigma}_\epsilon$	SBB 95% CI
Poisson	1.8121	(1.4710, 2.1640)
NB	1.0916	(0.9600, 1.2410)
ZIP	1.1094	(0.6741, 1.2104)
ZINB	1.0707	(0.9230, 1.219)
ZICMP	1.1737	(0.9840, 1.381)

Table 6 shows the estimated value of σ_ϵ for zero-inflated marginals and the ordinary Poisson and NB marginals. After estimating the latent process using (49), we draw SBB

95% CIs for σ_ϵ under each marginals. The results show that 1 is included in all the CIs except for the Poisson marginal, which suggests that choosing the Poisson leads to violation of the assumption. Figure 24 features the randomized quantile residuals in normal probability and autocorrelation plots of the copula-based ZIP, ZINB and ZICMP models. The normal probability plots suggest the randomized quantile residuals of these three models follow the normal distribution, and the autocorrelation plots indicate the absence of the serial dependence in the residuals. These findings suggest that the proposed models in this chapter fit the data adequately. Models with more complicated correlation structures such as $AR(2)$ and $ARMA(1,1)$ were also considered and fitted to the data with the same covariates. No significant improvements were found and thus we recommend using $AR(1)$. However, dropping the trend and seasonality covariates and running the models with only the dust haze covariate yields significant $AR(2)$ and $ARMA(1,1)$ dependence structures. Figure 25 shows the predicted values of the sandstorm counts from the three proposed models. The predicted values were calculated using the conditional expectation of Y_t given the past $Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1$. The plots indicate that our copula-based zero-inflated models adequately predict the injury counts.

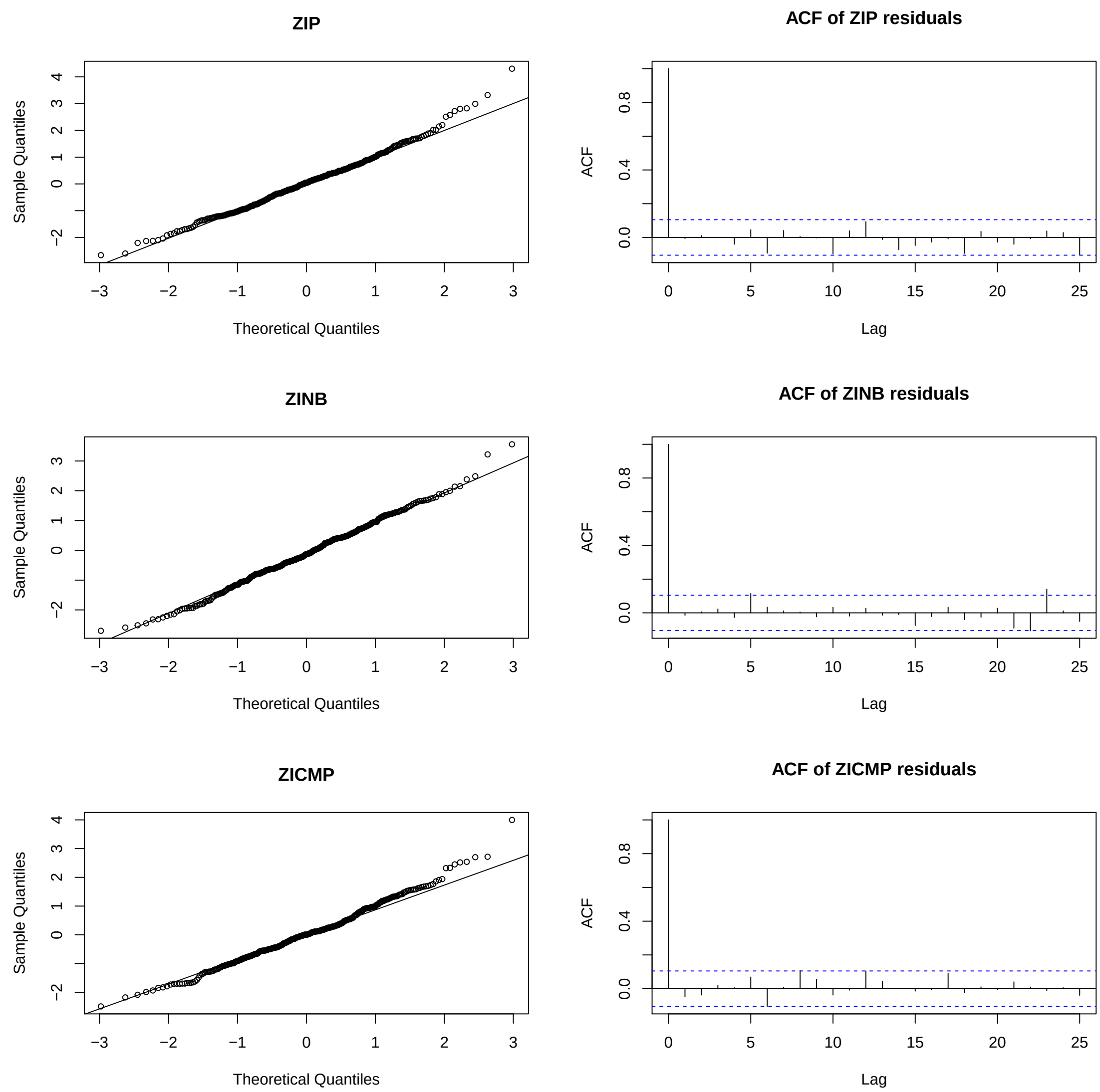


Figure 24: Sandstorm counts series: q-q plots (left) and autocorrelation plots (right) for sets of randomized residuals of the ZIP, ZINB and ZICMP models.

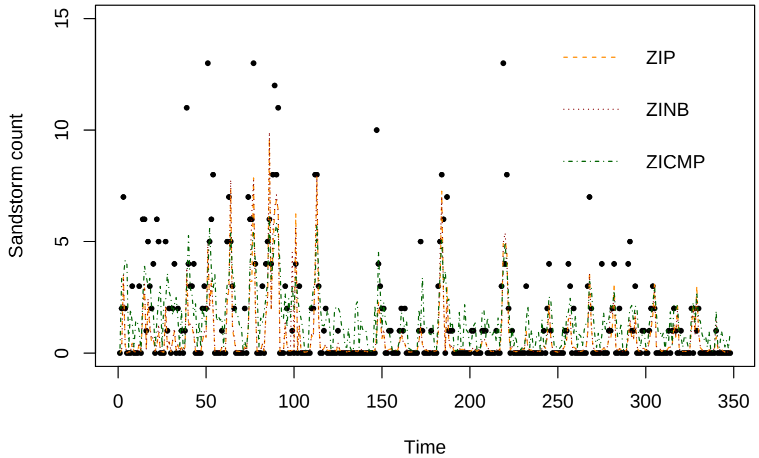


Figure 25: Prediction plot using the conditional expectations of the ZIP, ZINB and ZICMP models. Dots represent the observed sandstorm counts

CHAPTER 4

MARKOV ZERO-INFLATED COUNT TIME SERIES MODELS WITH COPULA-BASED TRANSITION PROBABILITIES

4.1 INTRODUCTION

This chapter concentrates on building a class of Markov zero-inflated count time series models based on a joint distribution on consecutive observations. The joint distribution function of the consecutive observations is constructed through copula functions. The Markov chains considered here are of first or second order with the univariate margins of ZIP, ZINB, or ZICMP distributions as defined in Section 2.2. Higher-order Markov models may be applied too by extending the work for second order models. However, zero-inflated count time series, and low counts in general, correspond to low order dependence structure quite often (Joe, 2016). Therefore, the work done in this chapter is mainly concerning first and second order Markov models.

For first order Markov models, bivariate copula functions such as the bivariate Gaussian, Frank, and Gumbel are chosen to construct a bivariate distribution of two consecutive observations. For second order Markov models, trivariate Gaussian copula function will be employed. In addition, other copula functions can be used to construct the trivariate joint distribution through suitable functions discussed later in this chapter.

The idea of constructing Markov chains with copula-based transition probabilities was first introduced in Chapter 8 in Joe (1997) as an application of copula theories. The advantage of using this model over the one introduced in Chapter 3 is that the $n-$dimensional multivariate distribution can be broken down to a function of $n-1$ bivariate or trivariate joint distributions. The bivariate or trivariate joint distributions are far more easier to handle than the $n-$dimensional joint distribution especially if n is large, and that is the case for most time series data.

Before discussing the proposed models, we will give brief review of Markov chains, the

big umbrella of the class of zero-inflated models we plan to present here. The theory of Markov chains is available in a great number of literature (see for examples, Cox and Miller (1965) and Serfozo (2009)). In discrete time series context, Raftery (1985) and later Adke and Deshmukh (1988) introduced a simple class of Markov chains models for a sequence of discrete time series variables, say $\{Y_t\}$, with values in a infinite, or finite, countable set S.

For first order Markov chains, consider the discrete-time stochastic process $\{Y_t\}$ (count time series) for $t = 1, \ldots, n$ on the infinite countable set $\mathcal{S} = \mathbb{Z}^+$, the set of nonnegative integers. The sequence $\{Y_t\}$ of $\mathcal{S}-$valued random variables is defined on a probability space $(\Omega, \mathcal{F}, P)$, where P is a probability measure on a $\sigma-$field $\mathcal{F}$ in an event-space Ω, and $\mathcal{S}$ is the state space of the process, and the count $Y_t \in \mathcal{S}$ is the state of the process at time t, for $t = 1, \ldots, n$. Now, the first order Markov chains is defined as follow.

Definition 3. *A stochastic process $\{Y_t\}$, for $t = 1, \ldots, n$ on a countable set $\mathcal{S}$ is a first order Markov chain if, for any $y_t, y_{t-1} \in \mathcal{S}$,*

$$P(Y_t = y_t | Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1) \quad = \quad P(Y_t = y_t | Y_{t-1} = y_{t-1}). \tag{52}$$

The right hand of (52) is the probability that the Markov chain jumps from state y_{t-1} to state y_t, and it is known as the *Markov property*. That is, at any time t, the next state y_{t+1} is conditionally independent of the past $y_1, y_2, \ldots, y_{t-1}$ given the present state y_t. For higher order Markov chains, Raftery (1985) introduced a class of simple models that accounts higher order dependence structure. In addition, the use of Markov chains was also found in count time series models based on thinning operators, as mentioned in Section 2.1.2. Another class that used Markov chains is found in constructing the count time series models based on multivariate distributions with random variables in a convolution-closed infinitely divisible class (Joe, 1996 and Jung and Tremayne, 2011). Here, we extend the work done by Joe (2014) on constructing the count time series models through copula-based joint distributions of consecutive observations. When the Markov process is continuous, the copula-based Markov models have been extensively examined by many (see for examples, Darsow et al., 1992, Joe, 1997, Chen and Fan, 2006, and Ibragimov, 2009).

The next sections detail the proposed class of Markov zero-inflated count time series models with copula-based transition probabilities. A parametric copula family is used for the joint distribution function of two consecutive counts in the case of first order Markov chain or three consecutive counts in the case of second order Markov chain. The use of

copula here allows to avoid some strict distributional assumptions on the marginals such as the infinite divisibility condition (Joe, 2016). The latter condition is not necessarily satisfied when we assume the counts follow a zero-inflated distribution. In addition, the copula-based models extend nicely to non-stationary processes through time-varying parameters in the univariate margins given in (9), (14), and (20), which is the goal of this chapter.

4.2 MARKOV CHAIN MODELS

A general form of first order Markov models with copula-based transition probabilities as defined in Joe (2014) is given by

$$Y_t = g(\epsilon_t; Y_{t-1}),$$

where $\{\epsilon_t\}$ is an i.i.d stochastic continuous latent process, and $g(.)$ is assumed to an increasing function in ϵ_t for $t = 1, \ldots, n$. Thus, the observed value Y_t depends on the past through only Y_{t-1}. If the process $\{Y_t\}$ is continuous, then there exists a simple stochastic representation for the Markov model. However, for for discrete process, as in the case here, there is no simple stochastic representation for the model (Joe, 2016).

4.2.1 FIRST ORDER MARKOV MODELS

Suppose $\{Y_t\}$ is zero-inflated count time series first order Markov chains following one of the distribution introduced in Section 2.2. Let f_t and F_t be the pdf and the cdf of Y_t, respectively. Then, taking advantage of the chain rule of probability and the Markov property, the multivariate joint density distribution of $Y_1, \ldots, Y_n$ is given by

$$\begin{aligned} \Pr(Y_1 = y_1, \ldots, Y_n = y_n) &= \prod_{t=1}^{n} \Pr(Y_t = y_t | Y_1 = y_1, \ldots, Y_{t-1} = y_{t-1}) \\ &= \Pr(Y_1 = Y_1 = y_1 y_1) \prod_{t=2}^{n} \Pr(Y_t = y_t | Y_{t-1} = y_{t-1}), \quad (53) \end{aligned}$$

the transition probability, i.e. conditional probability in the right hand of (53) depends on the joint density function of Y_t, Y_{t-1} and can be found using the copula functions as

introduced in Chapter 2. That is, let

$$F_{12}(y_t, y_{t-1}) = C(F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta}), F_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta});\boldsymbol{\delta}),$$

where $C(.;\boldsymbol{\delta})$ is a bivariate copula function with parameter vector $\boldsymbol{\delta}$. The covariates $\mathbf{X}_t = (\mathbf{x}_t, \mathbf{z}_t, \mathbf{w}_t)$, for $t = 1, \ldots, n$ are the covariates corresponding to the intensity (mean) parameter λ_t, the zero-inflation parameter ω_t and the dispersion parameter κ_t if existed, respectively. Notice that in some cases, these parameters or part of them are constant across the time when the covariates are not significant and dropped from the model. The parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$ is the unknown marginal regression coefficient. Hence, the transition probability is given by

$$\Pr(Y_t = y_t|Y_{t-1} = y_{t-1}) = \frac{\Pr(Y_t = y_t, Y_{t-1} = y_{t-1})}{f_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta})}, \tag{54}$$

where

$$\Pr(Y_t = y_t, Y_{t-1} = y_{t-1}) = F_{12}(y_t, y_{t-1}) - F_{12}(y_t^-, y_{t-1}) - F_{12}(y_t, y_{t-1}^-) + F_{12}(y_t^-, y_{t-1}^-),$$

and $y_t^- = y_t - 1$ since Y_t is a discrete random variable, for all t.

Several choices of the bivariate copula function $C(.;\boldsymbol{\delta})$ can be selected depending on the degree and the sign of the dependence and the tail behavior of the copula. For example, if there is symmetry, the Gaussian copula is recommended. However, Gumbel or reflected Gumbel copula perform better than Gaussian copula in the existence of tail dependence (Joe, 2016). Moreover, the Frank and the Gaussian copulas are both reflection symmetric and allow for negative dependence (see Chapter 2 for more details).

Next, we will discuss the immediate extension of the first order Markov models. That is, the second order Markov models where the zero-inflated count Y_t depended on the past two counts.

4.2.2 SECOND ORDER MARKOV MODELS

Suppose $\{Y_t\}$ is zero-inflated count time series of second order Markov chains following one of the distribution introduced in Section 2.2. Let f_t and F_t be the pdf and the cdf of the observed count Y_t, respectively. Now, the multivariate joint distribution of $Y_1, \ldots, Y_n$ is

given by

$$\begin{aligned}
\Pr(Y_1 = y_1, \ldots, Y_n = y_n) &= \prod_{t=1}^{n} \Pr(Y_t = y_t | Y_1 = y_1, \ldots, Y_{t-1} = y_{t-1}) \\
&= \Pr(Y_1 = y_1) \Pr(Y_2 = y_2 | Y_1 = y_1) \\
&\times \prod_{t=3}^{n} \Pr(Y_t = y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}) \\
&= \Pr(Y_1 = y_1, Y_2 = y_2) \\
&\times \prod_{t=3}^{n} \Pr(Y_t = y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}), \qquad (55)
\end{aligned}$$

where the conditional probability of Y_2 given $Y_1 = y_1$ can be evaluated using (54). However, for the second order transition probabilities of Y_t given $Y_{t-1} = y_{t-1}$ and $Y_{t-2} = y_{t-2}$, we need to fit an appropriate trivariate copula function for the joint distribution of Y_t, Y_{t-1} and Y_{t-2} for $t = 3, \ldots, n$.

The most popular choice is the trivariate Gaussian copula. Using the joint multivariate distribution with discrete margin given in (22) and the Gaussian copula function given in (23), the transition probabilities of Y_t given $Y_{t-1} = y_{t-1}$ and $Y_{t-2} = y_{t-2}$ for $t = 3, \ldots, n$ is given by

$$\Pr(Y_t = y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}) = \frac{\Pr(Y_t = y_t, Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2})}{\Pr(Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2})},$$

where the joint distribution is given by

$$\begin{aligned}
&\Pr(Y_t = y_t, Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}) = \\
&F_{123}(y_t, y_{t-1}, y_{t-2}) - F_{123}(y_t, y_{t-1}, y_{t-2}^-) - \\
&F_{123}(y_t, y_{t-1}^-, y_{t-2}) - F_{123}(y_t^-, y_{t-1}, y_{t-2}) + \\
&F_{123}(y_t, y_{t-1}^-, y_{t-2}^-) + F_{123}(y_t^-, y_{t-1}, y_{t-2}^-) +
\end{aligned}$$

$$F_{123}(y_t^-, y_{t-1}^-, y_{t-2}) + F_{123}(y_t^-, y_{t-1}^-, y_{t-2}^-), \tag{56}$$

where the function $F_{123}(.)$ is given by

$$F_{123}(y_t, y_{t-1}, y_{t-2}) =$$

$$\Phi_{R(\boldsymbol{\delta})}(F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta}), F_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta}), F_{t-2}(y_{t-2}|\mathbf{X}_{t-2};\boldsymbol{\theta})),$$

for $t = 3, \ldots, n$ where $R(\boldsymbol{\delta})$ is a 3×3 correlation matrix, with $\boldsymbol{\delta} = (\delta_1, \delta_2)'$ as a vector of the dependence structure parameters. The covariates $\mathbf{X}_t = (\mathbf{x}_t, \mathbf{z}_t, \mathbf{w}_t)$, for $t = 1, \ldots, n$ are the covariates corresponding to the intensity (mean) parameter λ_t, the zero-inflation parameter ω_t and the dispersion parameter κ_t if existed, respectively. The parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\alpha}')'$ is of the unknown marginal regression coefficient. The bivariate copula margins F_{12} and F_{23} of the trivariate copula function F_{123} are the same and given by the bivariate Gaussian copula in this case.

Another way of evaluating the trivariate joint density function, when the Gaussian copula function is chosen, can be through integration over rectangle probability. That is

$$\Pr(Y_t = y_t, Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}) =$$

$$\int_{\mathcal{D}_t(y_t;\boldsymbol{\theta})} \int_{\mathcal{D}_{t-1}(y_{t-1};\boldsymbol{\theta})} \int_{\mathcal{D}_{t-2}(y_{t-2};\boldsymbol{\theta})} \phi_{R(\boldsymbol{\delta})}(z_{t-2}, z_{t-1}, z_t) dz_{t-2} dz_{t-1} dz_t, \tag{57}$$

where

$$\mathcal{D}_t(y_t;\boldsymbol{\theta}) = [\Phi^{-1}\{F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta})\}, \Phi^{-1}\{F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta})\}]. \tag{58}$$

Although the Gaussian copula function in (57) has no closed form, there are several accurate deterministic approximations of the function when the dimension is low such as the case here with the trivariate Gaussian or the bivariate Gaussian. Further discussion on these approximation methods is on Section 4.3.

Another way of calculating the trivariate joint distribution, if the closed copula function

are desired, can be found by employing the Laplace transform (LT) of a non-negative random variable through a max-infinite divisible (max-id) copula given in (25). Joe and Hu (1996) stated that when the copula functions in (25) were chosen to be $C_{12} = C_{23} = H$, where H is a permutation symmetric max-id bivariate copula function, and C_{13} is the independent copula function with $v_1 = v_3 = 1$ and $v_2 = 0$, then the model would be appropriate for generating a second order Markov chain.

Hence, the function $F_{123}(.)$ in (56) becomes the following trivariate max-id copula

$$F_{123}(y_t, y_{t-1}, y_{t-2}) =$$

$$\psi\Bigg(\sum_{j\in\{t,t-2\}} \Big[-\log H(e^{-0.5\psi^{-1}(F_j;\delta_1)}, e^{-0.5\psi^{-1}(F_{t-1};\delta_1)}; \delta_2) + \frac{1}{2}\psi^{-1}(F_j;\delta_1)\Big]; \delta_1 \Bigg), \quad (59)$$

where $F_j = F_j(y_j|\mathbf{X}_j; \boldsymbol{\theta})$ for $j = t, t-2$ and $t = 3, \ldots, n$. The function $\psi(.; \delta_1)$ is the Laplace transform, and the function $H(.; \delta_2)$ is a permutation symmetric max-id bivariate copula function. The bivariate margins of (59) are given by

$$\begin{aligned} F_{i2}(y_j, y_{t-1}) = \psi\Bigg(&- \log H(e^{-0.5\psi^{-1}(F_j;\delta_1)}, e^{-0.5\psi^{-1}(F_{t-1};\delta_1)}; \delta_2) \\ &+ \frac{1}{2}\psi^{-1}(F_j;\delta_1) + \frac{1}{2}\psi^{-1}(F_{t-1};\delta_1); \delta_1\Bigg), \end{aligned} \quad (60)$$

for $i = 1, 2$, $j = t, t-2$, and

$$F_{13}(y_t, y_{t-2}) = \psi(\psi^{-1}(F_t; \delta_1) + \psi^{-1}(F_{t-2}; \delta_1); \delta_1).$$

The above trivariate max-id copula is suggested to be used when there is stronger dependence for measurements at nearer time points (Joe, 2016). He also stated that in the case of large value clustering (such as when the time series observes seasonality) a good choice for $H(.; \delta_2)$ is the bivariate Gumbel copula and $\psi(.; \delta_1)$ is the positive stable Laplace transform, which results in having the function $F_{123}(.)$ in (59) to be a trivariate extreme value copula (see Appendix A.2 for the full expression of such functions).

The Gaussian copula and the max-id copula can be extended to fit Markov models of order greater than 2. In the next subsection, we will state some of the properties and advantages of using copula-based transition probabilities instead of applying traditional methods such as the one based on thinning operators.

4.2.3 MODEL PROPERTIES

A summary of some of the properties and advantages of using copula-based transition probabilities are given in Joe (2016). For copula-based transition probabilities, the following properties hold.

- Any marginal distribution can be used. Hence, in our case the ZIP, ZINB, and ZICMP distributions are considered.
- The serial dependence can be positive or negative depending on the choice of the copula function, unlike the case based on thinning operator where only positive autocorrelation can be used.
- Covariates can be included in the model to fit non-stationary time series through the parameters of the univariate regression model.
- Extension from first order Markov model to second order (or even higher) Markov model can be easily obtained by employing the techniques presented in Section 4.2.2.
- More options of serial correlation functions are available than those based on thinning operators.
- Estimation via Likelihood inference can be easily applied especially if the copula family has a simple form.

4.2.4 MEASURE OF DEPENDENCE

When the chosen copula family is the Gaussian, interpreting the dependence parameter, δ, is straightforward since it corresponds to the Pearson's correlation. However, for other copula families, there is no clear meaning of the value of the dependence parameter δ. Each dependence, or copula, parameter family has different range depending on the copula family. Consequently, it is difficult to compare and interpret the values from different families. In order to compare the degree of dependence through the copula parameter δ, one can apply the copula-based Kendall's tau correlation coefficient. If the margins are continuous, the value of Kendall's tau depends only on the parameter δ (Nelsen, 2007). However, for discrete margins, as of the case here, the value of Kendall's tau depends on the choice of the

marginal distributions too since there is a chance of a tie when deriving the concordance probability. One can use the value of Kendall's tau that depends only on the parameter δ but then it would account only on the dependence dominated by the middle of the data (Nikoloulopoulos and Moffatt, 2019). Nikoloulopoulos and Karlis (2010) derived a formula for Kendall's tau for discrete margins. Through their formula, they showed that the value of Kendall's tau was affected by the choice of the marginal distribution. For normalized versions of Kendall's tau, one can refer to Goodman and Kruskal (1954) and Nešlehová (2007). In this chapter, we compare between the dependence of different copula families using the value of Kendall's tau given in Nelsen (2007).

4.3 STATISTICAL INFERENCE

4.3.1 LOG-LIKELIHOOD FUNCTIONS

Inference or estimation method performed for the Markov models' parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \boldsymbol{\delta}')'$ presented in this chapter is the maximum likelihood estimation (MLE) method. As stated before likelihood inference method is easily applied when the chosen copula family has simple form. In addition, likelihood inference gives us the advantage of performing hypothesis testing through the likelihood ratio statistics and model selection through the log-likelihood function. Next, we give a detailed description of the likelihood functions of the two models presented earlier.

For the first order Markov models, the likelihood function is given by

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

with the log-likelihood given as

$$\begin{aligned} l(\boldsymbol{\vartheta}; \mathbf{y}) &= \log \Pr(Y_1 = y_1; \boldsymbol{\theta}) + \sum_{t=2}^{n} \log \Pr(Y_t = y_t | Y_{t-1} = y_{t-1}; \boldsymbol{\vartheta}) \\ &= \log f_1(y_1 | \mathbf{X}_1; \boldsymbol{\theta}) + \sum_{t=2}^{n} \log \frac{\Pr(Y_t = y_t, Y_{t-1} = y_{t-1}; \boldsymbol{\vartheta})}{f_{t-1}(y_{t-1} | \mathbf{X}_{t-1}; \boldsymbol{\theta})} \\ &= \log f_1(y_1 | \mathbf{X}_1; \boldsymbol{\theta}) \end{aligned}$$

$$
\begin{aligned}
&+ \sum_{t=2}^{n} \Big[\log\{C(F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta}), F_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta});\boldsymbol{\delta}) \\
&- C(F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta}), F_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta});\boldsymbol{\delta}) \\
&- C(F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta}), F_{t-1}(y_{t-1}^-|\mathbf{X}_{t-1};\boldsymbol{\theta});\boldsymbol{\delta}) \\
&+ C(F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta}), F_{t-1}(y_{t-1}^-|\mathbf{X}_{t-1};\boldsymbol{\theta});\boldsymbol{\delta})\} \\
&- \log f_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta})\Big], \qquad (61)
\end{aligned}
$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ are the parameter vectors of the marginals and the dependence structure, respectively. The log-likelihood function in (61) has closed form if the copula family chosen to define $C(.;\boldsymbol{\delta})$ has a closed form. 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. That is,

$$
C(F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta}), F_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta});\delta) =
$$

$$
\int_{-\infty}^{\mathcal{D}_t^+} \int_{-\infty}^{\mathcal{D}_{t-1}^+} \phi_{R(\delta)}(z_t, z_{t-1}) dz_{t-1} dz_t,
$$

where $\mathcal{D}_t^+ = \Phi^{-1}(F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta}))$ and $\mathcal{D}_{t-1}^+ = \Phi^{-1}(F_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta}))$, for $t = 2, \ldots, n$ and the 2×2 correlation matrix $R(\delta)$ is given by

$$
R(\delta) = \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix}.
$$

However, accurate deterministic approximation methods of the normal probability integrals are available for low dimensions on standard softwares, see Hothorn et al. (2001) for example. In contrast, choosing Frank copula family, for instance, will lead to a closed form likelihood function that can be maximized easily. The copula function in this case will be given by

$$
C(F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta}), F_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta});\delta) =
$$

$$-\frac{1}{\delta}\log\left[1-\frac{(1-e^{-\delta F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta})})(1-e^{-\delta F_{t-1}(y_{t-1}|\mathbf{X}_{t-1};\boldsymbol{\theta})})}{1-e^{-\delta}}\right],$$

for $t = 2, \ldots, n$, which obviously has a closed form and hence the log-likelihood function given in (61) has a closed form. Consequently, maximizing the function would be much faster. Other copula families also provide a closed form log-likelihood function such as the Gumbel and the reflected Gumbel copulas.

For second order Markov models, a general form of the log-likelihood function is given by

$$\begin{aligned} l(\boldsymbol{\vartheta};\mathbf{y}) &= \log \Pr(Y_1 = y_1, Y_2 = y_2) \sum_{t=3}^{n} \log \Pr(Y_t = y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}) \\ &= \log \Pr(Y_1 = y_1, Y_2 = y_2) \\ &+ \sum_{t=3}^{n} \Big[\log \Pr(Y_t = y_t, Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}) \\ &- \log \Pr(Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}) \Big], \end{aligned} \tag{62}$$

where the bivariate and trivariate joint densities depend on the choice of the copula construction methods presented in Section 4.2.2. When the Gaussian copula function is chosen, the log-likelihood function in (62) can be given by

$$\begin{aligned} l(\boldsymbol{\vartheta};\mathbf{y}) &= \log\Big\{ \int_{\mathcal{D}_1(y_1;\boldsymbol{\theta})} \int_{\mathcal{D}_2(y_2;\boldsymbol{\theta})} \phi_{R(\boldsymbol{\delta})}(z_2, z_1) dz_2 dz_1 \Big\} \\ &+ \sum_{t=3}^{n} \Big[\log\Big\{ \int_{\mathcal{D}_t(y_t;\boldsymbol{\theta})} \int_{\mathcal{D}_{t-1}(y_{t-1};\boldsymbol{\theta})} \int_{\mathcal{D}_{t-2}(y_{t-2};\boldsymbol{\theta})} \phi_{R(\boldsymbol{\delta})}(z_{t-2}, z_{t-1}, z_t) dz_{t-2} dz_{t-1} dz_t \Big\} \\ &- \log\Big\{ \int_{\mathcal{D}_{t-1}(y_{t-1};\boldsymbol{\theta})} \int_{\mathcal{D}_{t-2}(y_{t-2};\boldsymbol{\theta})} \phi_{R(\boldsymbol{\delta})}(z_{t-2}, z_{t-1}) dz_{t-2} dz_{t-1} \Big\} \Big], \end{aligned} \tag{63}$$

where $\mathcal{D}_t(y_t;\boldsymbol{\theta})$ is given by (58). The expression in (63) is not in a closed form, and approximations are needed for the rectangle probabilities. However, for the trivariate max-id copula, the log-likelihood function can take a closed form and is given by

$$l(\boldsymbol{\vartheta};\mathbf{y}) = \log\Big\{ \sum_{j_1=0}^{1} \sum_{j_2=0}^{1} F_{12}(y_1 - j_1, y_2 - j_2) \Big\}$$

$$+ \sum_{t=3}^{n} \Big[\log \Big\{ \sum_{j_1=0}^{1} \sum_{j_2=0}^{1} \sum_{j_3=0}^{1} (-1)^{j_1+j_2+j_3} F_{123}(y_t - j_1, y_{t-1} - j_2, y_{t-2} - j_3) \Big\}$$
$$- \log \Big\{ \sum_{j_1=0}^{1} \sum_{j_2=0}^{1} F_{12}(y_{t-1} - j_1, y_{t-2} - j_2) \Big\} \Big], \quad (64)$$

where $F_{12}(y_{t-1}, y_{t-2})$ and $F_{123}(y_t, y_{t-1}, y_{t-2})$ are given by (60) and (59), respectively.

Hence, the maximum likelihood estimates of $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \boldsymbol{\delta}')'$ can be obtained by:

$$\widehat{\boldsymbol{\vartheta}} = \arg\max_{\boldsymbol{\vartheta}} \; l(\boldsymbol{\vartheta}; \mathbf{y}).$$

This optimization will produce a Hessian Matrix that yields the observed Fisher information matrix. To get the standard errors of the ML estimates of $\boldsymbol{\vartheta}$, one can take the inverse of the Fisher information matrix. In the next sections, score functions are derived and asymptotic results are derived to prove that the inverse of the Fisher information matrix of the likelihood functions in (61) and (62) evaluated at the MLE of $\boldsymbol{\vartheta}$ can be used as an estimated covariance of matrix of $\widehat{\boldsymbol{\vartheta}}$.

4.3.2 SCORE FUNCTIONS

The score functions of the log-likelihood functions given in (61), (63), and (64) can be obtained from the following general definition of the function.

$$\begin{aligned} S(\boldsymbol{\vartheta}) &= \frac{\partial l(\boldsymbol{\vartheta}; \mathbf{y})}{\partial \boldsymbol{\vartheta}} \\ &= \left[\frac{\partial l(\boldsymbol{\vartheta}; \mathbf{y})}{\partial \boldsymbol{\theta}}, \frac{\partial l(\boldsymbol{\vartheta}; \mathbf{y})}{\partial \boldsymbol{\delta}} \right], \end{aligned} \quad (65)$$

where $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha})'$ is a vector of the marginal parameters and $\boldsymbol{\delta}$ is a vector of the dependence (or copula) parameters. The function in (65) requires the derivation of the marginal pdfs and cdfs of the ZIP, ZINB, and ZICMP distributions alongside the copula functions.

In particular, consider the log-likelihood function of the first order Markov process

given in (61). Thus, (65) is given by

$$S(\boldsymbol{\vartheta}) = \left[\frac{\partial l(\boldsymbol{\vartheta};\mathbf{y})}{\partial \boldsymbol{\gamma}}, \frac{\partial l(\boldsymbol{\vartheta};\mathbf{y})}{\partial \boldsymbol{\beta}}, \frac{\partial l(\boldsymbol{\vartheta};\mathbf{y})}{\partial \boldsymbol{\alpha}}, \frac{\partial l(\boldsymbol{\vartheta};\mathbf{y})}{\partial \delta}\right], \tag{66}$$

where $\frac{\partial l(\boldsymbol{\vartheta};\mathbf{y})}{\partial \boldsymbol{\alpha}}$ is omitted when the ZIP distribution is considered. The components of (66) for the marginal parameters are given by

$$\frac{\partial l(\boldsymbol{\vartheta};\mathbf{y})}{\partial \theta} = \frac{f'_{1,\theta}}{f_1} + \sum_{t=2}^{n}\left[\frac{\Delta C'_{t,\theta}}{\Delta C_t} - \frac{f'_{t,\theta}}{f_t}\right],$$

where θ can be any marginal parameter, f_t is the marginal pdf and $f'_{t,\theta}$ is its corresponding partial derivative with respect to the marginal parameter θ,

$$\begin{aligned}\Delta C_t &= C(F_t, F_{t-1};\delta) - C(F_t^-, F_{t-1};\delta) \\ &- C(F_t, F_{t-1}^-;\delta) + C(F_t, F_{t-1};\delta),\end{aligned}$$

$\Delta C'_{t,\theta}$ is its corresponding partial derivative with respect to the marginal parameter θ, F_t is the marginal cdf, $F'_{t,\theta}$ is its corresponding partial derivative with respect to the marginal parameter θ, and $F_t^- = F_t(y^-)$. For the dependence parameter δ, the partial derivative is given by

$$\frac{\partial l(\boldsymbol{\vartheta};\mathbf{y})}{\partial \delta} = \sum_{t=2}^{n}\frac{\Delta C'_{t,\delta}}{\Delta C_t}.$$

The partial derivative of the copula function with respect to the marginal parameter θ can be obtain using the chain rule, that is if $\theta = \gamma_i$ a regression coefficient of a covariate associated with the zero-inflated parameter ω_t for $i = 1, \ldots, l$ and $t = 1, \ldots, n$, then

$$\frac{\partial C}{\partial \gamma_i} = \frac{\partial C}{\partial F_t}\frac{\partial F_t}{\partial \omega_t}\frac{\partial \omega_t}{\partial \gamma_i} + \frac{\partial C}{\partial F_{t-1}}\frac{\partial F_{t-1}}{\partial \omega_{t-1}}\frac{\partial \omega_{t-1}}{\partial \gamma_i},$$

if $\theta = \beta_i$ a regression coefficient of a covariate associated with the intensity parameter λ_t for $i = 1, \ldots, k$ then

$$\frac{\partial C}{\partial \beta_i} = \frac{\partial C}{\partial F_t}\frac{\partial F_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \beta_i} + \frac{\partial C}{\partial F_{t-1}}\frac{\partial F_{t-1}}{\partial \lambda_{t-1}}\frac{\partial \lambda_{t-1}}{\partial \beta_i},$$

and finally if $\theta = \alpha_i$ a regression coefficient of a covariate associated with the dispersion parameter (if existed) κ_t for $i = 1, \ldots, m$ then

$$\frac{\partial C}{\partial \alpha_i} = \frac{\partial C}{\partial F_t}\frac{\partial F_t}{\partial \kappa_t}\frac{\partial \kappa_t}{\partial \alpha_i} + \frac{\partial C}{\partial F_{t-1}}\frac{\partial F_{t-1}}{\partial \kappa_{t-1}}\frac{\partial \kappa_{t-1}}{\partial \alpha_i}.$$

The partial derivative with respect to the dependence parameter δ is basically given by

$$\frac{\partial C}{\partial \delta}.$$

The partial derivatives of the link functions are given as follow,

$$\frac{\partial \omega_t}{\partial \gamma_i} = \frac{z_{it} e^{\boldsymbol{\gamma}'\mathbf{z}_t}}{(1 + e^{\boldsymbol{\gamma}'\mathbf{z}_t})^2},$$

$$\frac{\partial \lambda_t}{\partial \beta_i} = x_{it} e^{\boldsymbol{\beta}'\mathbf{x}_t},$$

and

$$\frac{\partial \kappa_t}{\partial \beta_i} = w_{it} e^{\boldsymbol{\alpha}'\mathbf{w}_t}.$$

The following are the partial derivatives of the marginal distributions with respect to ω_t, λ_t, and κ_t (if existed). For the ZIP distribution, the partial derivatives of the density function given in (6) are

$$\frac{\partial f_t(y_t)}{\partial \omega_t} = y_{0,t} - \frac{e^{-\lambda_t}\lambda_t^{y_t}}{y_t!},$$

where $y_{0,t}$ is 1 if $y_t = 0$ and 0 otherwise, and

$$\begin{aligned} \frac{\partial f_t(y_t)}{\partial \lambda_t} &= (1 - \omega_t)\frac{1}{y_t!}\left[e^{-\lambda_t} y_t \lambda_t^{y_t - 1} - e^{-\lambda_t}\lambda_t^{y_t}\right] \\ &= (1 - \omega_t)\frac{e^{-\lambda_t}\lambda_t^{y_t - 1}(y_t - \lambda_t)}{y_t!}. \end{aligned}$$

The partial derivatives of the cdf of the ZIP distribution given in (9) are

$$\frac{\partial F_t(y_t)}{\partial \omega_t} = 1 - e^{-\lambda_t}\sum_{m=0}^{y_t}\frac{\lambda_t^m}{m!},$$

and

$$\begin{aligned}
\frac{\partial F_t(y_t)}{\partial \lambda_t} &= (1-\omega_t)\Bigg[e^{-\lambda_t}\sum_{m=0}^{y_t} m\frac{\lambda_t^{m-1}}{m!} - e^{-\lambda_t}\sum_{m=0}^{y_t}\frac{\lambda_t^m}{m!}\Bigg] \\
&= (1-\omega_t)e^{-\lambda_t}\Bigg[\sum_{m=1}^{y_t}\frac{\lambda_t^{m-1}}{(m-1)!} - \sum_{m=0}^{y_t}\frac{\lambda_t^m}{m!}\Bigg] \\
&= -(1-\omega_t)e^{-\lambda_t}\frac{\lambda_t^{y_t}}{y_t!}.
\end{aligned}$$

For the ZINB distribution, the marginal density given in (13) can be expressed as

$$f_t(y_t) = \omega_t y_{0,t} + (1-\omega_t)\exp\{Q_1\},$$

where

$$Q_1 = \log(\Gamma(\kappa_t+y_t)) - \log(\Gamma(\kappa_t)) + \kappa_t\log(\kappa_t) - \kappa_t\log(\kappa_t+\lambda_t) + y_t\log(\lambda_t) - y_t\log(\kappa_t+\lambda_t).$$

Hence,

$$\frac{\partial f_t(y_t)}{\partial \omega_t} = y_{0,t} - \exp\{Q_1\},$$

$$\begin{aligned}
\frac{\partial f_t(y_t)}{\partial \lambda_t} &= (1-\omega_t)\Bigg[-\frac{\kappa_t}{\kappa_t+\lambda_t} + \frac{y_t}{\lambda_t} - \frac{y_t}{\kappa_t+\lambda_t}\Bigg]\exp\{Q_1\} \\
&= \frac{\kappa_t(1-\omega_t)(y_t-\lambda_t)}{\lambda_t(\kappa_t+\lambda_t)}\exp\{Q_1\},
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial f_t(y_t)}{\partial \kappa_t} &= (1-\omega_t)\Bigg[\frac{\Gamma'(\kappa_t+y_t)}{\Gamma(\kappa_t+y_t)} - \frac{\Gamma'(\kappa_t)}{\Gamma(\kappa_t)} + 1 + \log(\kappa_t) \\
&- \frac{\kappa_t}{\kappa_t+\lambda_t} - \log(\kappa_t+\lambda_t) - \frac{y_t}{\kappa_t+\lambda_t}\Bigg]\exp\{Q_1\} \\
&= (1-\omega_t)\Bigg[1 + \psi(\kappa_t+y_t) - \psi(\kappa_t) + \log\left(\frac{\kappa_t+\lambda_t}{\kappa_t}\right) - \frac{\kappa_t+y_t}{\kappa_t+\lambda_t}\Bigg]\exp\{Q_1\},
\end{aligned}$$

where $\psi(.) = \frac{\Gamma'(.)}{\Gamma(.)}$ is the digamma function.

The partial derivatives of the cdf of the ZINB distribution are obtained after expressing the

cdf given in (14) as

$$F_t(y_t) = \omega_t + (1-\omega_t)\exp\{Q_2\}$$

where

$$Q_2 = \kappa_t \log(\kappa_t) - \kappa_t \log(\kappa_t+\lambda_t) - \log(\Gamma(\kappa_t)) + \log \sum_{m=0}^{y_t} e^{q_m},$$

and

$$q_m = \log(\Gamma(\kappa_t+m)) - \log(m!\,) + m\log(\lambda_t) - m\log(\kappa_t+\lambda_t).$$

Hence,

$$\frac{\partial F_t(y_t)}{\partial \omega_t} = 1-\exp\{Q_2\},$$

$$\begin{aligned}
\frac{\partial F_t(y_t)}{\partial \lambda_t} &= (1-\omega_t)\left[-\frac{\kappa_t}{\kappa_t+\lambda_t} + \frac{\sum_{m=0}^{y_t}\left(\frac{m}{\lambda_t}-\frac{m}{\kappa_t+\lambda_t}\right)e^{q_m}}{\sum_{m=0}^{y_t} e^{q_m}}\right]\exp\{Q_2\} \\
&= (1-\omega_t)\frac{\kappa_t}{\kappa_t+\lambda_t}\left[\frac{1}{\lambda_t}\frac{\sum_{m=0}^{y_t} m e^{q_m}}{\sum_{m=0}^{y_t} e^{q_m}} - 1\right]\exp\{Q_2\},
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial F_t(y_t)}{\partial \kappa_t} &= (1-\omega_t)\Bigg[1+\log(\kappa_t) - \frac{\kappa_t}{\kappa_t+\lambda_t} - \log(\kappa_t+\lambda_t) - \psi(\kappa_t) \\
&+ \frac{\sum_{m=0}^{y_t}\left(\psi(\kappa_t+m)\frac{m}{\kappa_t+\lambda_t}e^{q_m}\right)}{\sum_{m=0}^{y_t} e^{q_m}}\Bigg]\exp\{Q_2\} \\
&= (1-\omega_t)\Bigg[1+\log\left(\frac{\kappa_t}{\kappa_t+\lambda_t}\right)\frac{\kappa_t}{\kappa_t+\lambda_t} - \psi(\kappa_t) \\
&+ \frac{1}{\kappa_t+\lambda_t}\frac{\sum_{m=0}^{y_t} m\psi(\kappa_t+m)e^{q_m}}{\sum_{m=0}^{y_t} e^{q_m}}\Bigg]\exp\{Q_2\}.
\end{aligned}$$

For the ZICMP distribution, the marginal density given in (19) can be expressed as

$$f_t(y_t) = \omega_t y_{0,t} + (1-\omega_t)\exp\{Q_3\},$$

where

$$Q_3 = y_t \log(\lambda_t) - \kappa_t \log(y!\,) - Z,$$

where $Z = Z(\lambda_t, \kappa_t)$. Hence,

$$\frac{\partial f_t(y_t)}{\partial \omega_t} = y_{0,t} - \exp\{Q_3\},$$

$$\frac{\partial f_t(y_t)}{\partial \lambda_t} = (1-\omega_t)\Big[\frac{y_t}{\lambda_t} - \frac{\partial Z}{\partial \lambda_t}\frac{1}{Z}\Big]\exp\{Q_3\},$$

and

$$\frac{\partial f_t(y_t)}{\partial \kappa_t} = (1-\omega_t)\Big[-\frac{\kappa_t}{y_t!} - \frac{\partial Z}{\partial \kappa_t}\frac{1}{Z}\Big]\exp\{Q_3\}.$$

The partial derivatives of the cdf of the ZICMP distribution are obtained after expressing the cdf given in (20) as

$$F(y_t) = \omega_t + (1-\omega_t)\exp\{Q_4\}$$

where

$$Q_4 = \log \sum_{m=0}^{y_t} (m!\,)^{-\kappa_t}\lambda_t^m - \log(Z).$$

Hence,

$$\frac{\partial F_t(y_t)}{\partial \omega_t} = 1 - \exp\{Q_4\},$$

$$\frac{\partial F_t(y_t)}{\partial \lambda_t} = (1-\omega_t)\left[\frac{\sum_{m=1}^{y_t} m(m!\,)^{-\kappa_t}\lambda_t^{m-1}}{\sum_{m=0}^{y_t}(m!\,)^{-\kappa_t}\lambda_t^m} - \frac{\partial Z}{\partial \lambda_t}\frac{1}{Z}\right]\exp\{Q_4\},$$

and

$$\frac{\partial F_t(y_t)}{\partial \kappa_t} = -(1-\omega_t)\left[\frac{\sum_{m=0}^{y_t}(m!\,)^{-(\kappa_t+1)}\lambda_t^m}{\sum_{m=0}^{y_t}(m!\,)^{-\kappa_t}\lambda_t^m} + \frac{\partial Z}{\partial \kappa_t}\frac{1}{Z}\right]\exp\{Q_4\}.$$

For a bivariate copula function, $C(F_t, F_{t-1}; \delta)$, the partial derivative with respect to the marginal distribution function, F_t, is the conditional copula function F_{t-1} given F_t, and

vice versa. That is

$$\frac{\partial C}{\partial F_t} = C(F_{t-1}|F_t;\delta),$$

$$\frac{\partial C}{\partial F_{t-1}} = C(F_t|F_{t-1};\delta).$$

For example, if the copula function is chosen to be the Gaussian copula, then

$$\frac{\partial C}{\partial F_t} = \Phi\Big(\frac{\Phi^{-1}(F_{t-1}) - \delta\Phi^{-1}(F_t)}{\sqrt{1-\delta^2}}\Big),$$

$$\frac{\partial C}{\partial F_{t-1}} = \Phi\Big(\frac{\Phi^{-1}(F_t) - \delta\Phi^{-1}(F_{t-1})}{\sqrt{1-\delta^2}}\Big),$$

and

$$\frac{\partial C}{\partial \delta} = \phi(\Phi^{-1}(F_t), \Phi^{-1}(F_{t-1});\delta).$$

For the bivariate Frank copula,

$$\frac{\partial C}{\partial F_t} = e^{-\delta F_t}[(1-e^{-\delta})(1-e^{-\delta F_{t-1}})^{-1} - (1-e^{-\delta F_t})]^{-1},$$

$$\frac{\partial C}{\partial F_{t-1}} = e^{-\delta F_{t-1}}[(1-e^{-\delta})(1-e^{-\delta F_t})^{-1} - (1-e^{-\delta F_{t-1}})]^{-1},$$

and

$$\begin{aligned}\frac{\partial C}{\partial \delta} &= \frac{1}{\delta}\left[\frac{e^{-\delta} - F_{t-1}e^{-\delta F_{t-1}}(1-e^{-\delta F_t}) - F_t e^{-\delta F_t}(1-e^{-\delta F_{t-1}})}{(1-e^{-\delta}) - (1-e^{-\delta F_t})(1-e^{-\delta F_{t-1}})}\right] \\ &+ \frac{1}{\delta^2}\log\left[\frac{1-e^{-\delta}}{(1-e^{-\delta}) - (1-e^{-\delta F_t})(1-e^{-\delta F_{t-1}})}\right]\end{aligned}$$

One can also obtain the ML estimates of $\boldsymbol{\vartheta}$ through solving the score function, that is

$$S(\boldsymbol{\vartheta}) = 0.$$

4.3.3 ASYMPTOTIC PROPERTIES

To draw some inference on Markov processes, Billingsley (1961) gave important results that basically state that, under certain regularity conditions, the asymptotic likelihood

theory and numerical maximum likelihood from the i.i.d case can be extended to hold with dependent data following Markov process. First, we will consider the first order Markov model with the corresponding log-likelihood given in (61). As listed in Joe (1997), the regularity conditions needed to hold to use the results from Billingsley (1961) for the first order Markov process are verified. Before we list the regularity conditions, assume that we have $\{Y_t\}$ for $t = 1, \ldots, n$ a first order Markov chain with state space $\mathcal{S}$, and $\Pr(Y_t = y_t | Y_{t-1} = y_{t-1}; \boldsymbol{\vartheta})$ is a family of transition densities with respect to a counting measure and with column vector parameter $\boldsymbol{\vartheta}$ of dimension r in the parameter space $\boldsymbol{\Theta}$. In addition, for asymptotic analysis, rewrite the log-likelihood function given in (61) as

$$l_n(\boldsymbol{\vartheta}) = \sum_{t=2}^{n} p(\boldsymbol{\vartheta}; y_{t-1}, y_t), \tag{67}$$

where

$$p(\boldsymbol{\vartheta}; y_{t-1}, y_t) = \log \Pr(Y_t = y_t | Y_{t-1} = y_{t-1}; \boldsymbol{\vartheta}),$$

for $t = 2, \ldots, n$. Note that the first probability, $\Pr(Y_1 = y_1; \boldsymbol{\theta})$ is omitted from the function since the first observation, y_1, is asymptotically insignificant.

Regularity conditions are as follow.

(a) The maximum likelihood estimate $\widehat{\boldsymbol{\vartheta}}$ of $l_n(\boldsymbol{\vartheta})$ is assumed to satisfy

$$\frac{\partial}{\partial \boldsymbol{\vartheta}} l_n(\boldsymbol{\vartheta}) = \mathbf{0}.$$

(b) All states of the Markov chain communicate with each other (meaning that there are no transient states).

(c) The set of y for which $\Pr(y|x; \boldsymbol{\vartheta})$ is positive does not depend on $\boldsymbol{\vartheta}$.

(d) Let $\frac{\partial p}{\partial \boldsymbol{\vartheta}}$ be the column vector of partial derivatives $p_i = \partial p / \partial \vartheta_i$, $\frac{\partial^2 p}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'}$ be the matrix of second-order partial derivatives with components denoted by p_{ij}, and the third-order (mixed) derivatives are denoted by p_{ijk} for $i, j, k = 1, \ldots, r$. Then, p_i, p_{ij}, and p_{ijk} for $i, j, k = 1, \ldots, r$ exist and are continuous in $\boldsymbol{\vartheta}$.

(e) For $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, there exists a neighborhood $N_{\boldsymbol{\vartheta}}$ of $\boldsymbol{\vartheta}$ such that for all i, j, k,

$$\mathrm{E}_{\boldsymbol{\vartheta}} \Big[\sup_{\boldsymbol{\vartheta}' \in N_{\boldsymbol{\vartheta}}} |p_{ijk}(\boldsymbol{\vartheta}', Y_1, Y_2)| \Big] < \infty.$$

where $\mathrm{E}_{\boldsymbol{\vartheta}}$ means expectation assuming that the true parameter value is $\boldsymbol{\vartheta}$ and Y_1 start with a stationary distribution.

(f) For $i, j = 1, \ldots, r$,

$$\mathrm{E}_{\boldsymbol{\vartheta}}[|p_i(\boldsymbol{\vartheta}, Y_1, Y_2)|^2] < \infty,$$

and $\Sigma(\boldsymbol{\vartheta}) = (\sigma_{ij}(\boldsymbol{\vartheta}))$ is a non-singular $r \times r$ matrix with

$$\sigma_{ij}(\boldsymbol{\vartheta}) = \mathrm{E}_{\boldsymbol{\vartheta}}[p_i(\boldsymbol{\vartheta}, Y_1, Y_2), p_j(\boldsymbol{\vartheta}, Y_1, Y_2)].$$

Now, given these regularity conditions, the following asymptotic results from the i.i.d case hold form our Markov processes (Billingsley, 1961). In particular, there exists a root $\widehat{\boldsymbol{\vartheta}}_n$ of

$$\frac{\partial}{\partial \boldsymbol{\vartheta}} l_n(\boldsymbol{\vartheta}) = \mathbf{0}$$

such that

1. The ML estimator $\widehat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \boldsymbol{\delta}')'$ converges in probability to the true value of $\boldsymbol{\vartheta}$, say $\boldsymbol{\vartheta}^0$. That is, $\widehat{\boldsymbol{\vartheta}}_n$ is a consistent estimator of $\boldsymbol{\vartheta}$.

2. The ML estimator $\widehat{\boldsymbol{\vartheta}}_n$ is asymptotically normal. That is,

$$n^{1/2}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^0) \xrightarrow{d} N_n(\mathbf{0}, \Sigma^{-1}(\boldsymbol{\vartheta}^0)).$$

3. The log-likelihood ratios for hypotheses involving nested models for the parameter $\boldsymbol{\vartheta}$ have asymptotic chi-square distribution. That is,

$$2[\max_{\boldsymbol{\vartheta}} l_n(\boldsymbol{\vartheta}) - L(\boldsymbol{\vartheta}^0)] \xrightarrow{d} \chi^2_r.$$

Hence, as in the i.i.d case, the numerical maximization of (67) yields the observed Fisher information matrix which can be used as an estimated covariance matrix of $\widehat{\boldsymbol{\vartheta}}$. That is, for large n,

$$n^{-1}\Sigma^{-1}(\widehat{\boldsymbol{\vartheta}}) \approx I_n^{-1}(\widehat{\boldsymbol{\vartheta}}),$$

where

$$I_n^{-1}(\widehat{\boldsymbol{\vartheta}}) = -\Big[\frac{\partial^2 l_n(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'}\Big|_{\widehat{\boldsymbol{\vartheta}}}\Big]^{-1} = -\Big[\frac{\partial^2 \sum_{t=2}^n p(\boldsymbol{\vartheta}; y_{t-1}, y_t)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'}\Big|_{\widehat{\boldsymbol{\vartheta}}}\Big]^{-1}.$$

Joe (1997) argued that the theory in Billingsley (1961) still applies for higher-order Markov processes, assuming the order is known. He also stated that extension of the asymptotic theory to the case where the transition probabilities depend on covariates should be possible.

4.4 MODEL SELECTION AND PREDICTION

Choices of models should depend on their goodness of fit and predictive performances. There are several tools proposed to assess the models fit. It is advised to try more than one statistical model and compare results. One of the method used to compare fits of different models is the Akaike (Akaike, 1974) information criterion (AIC), which is given by

$$\mathrm{AIC}_1 = -2l(\widehat{\boldsymbol{\vartheta}}; \mathbf{y}) + 2r,$$

where $\widehat{\boldsymbol{\vartheta}}$ is the MLE and r is the number of parameter in a model. Differently, one might use the AIC as a penalized log-likelihood with the penalty being the number of parameters in a model (Joe, 1997). That is,

$$\mathrm{AIC}_2 = l(\widehat{\boldsymbol{\vartheta}}; \mathbf{y}) - r.$$

Many authors used these AICs to perform model selection for copula-based models (see for examples, Dias et al., 2004 and Palaro and Hotta, 2006). However, for larger sample size, the Bayesian information criterion (BIC) tends to provide better measure of fit (Shumway and Stoffer, 2011). The BIC is given by

$$\mathrm{BIC} = -2l(\widehat{\boldsymbol{\vartheta}}; \mathbf{y}) + r \log n.$$

Another way of comparing models performances is to evaluate their predictive performances. To measure predictive performance of a model, one might calculate the root mean square prediction error (RMSPE), and then pick the model with minimal RMSPE. The RMSPE for Markov models is defined in Joe (2014) as

$$\mathrm{RMSPE} = \left\{ \frac{1}{n-p} \sum_{t=1+p}^{n} \left[y_t - \widehat{\mathrm{E}}(Y_t | Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p}; \widehat{\boldsymbol{\vartheta}}) \right]^2 \right\}^{\frac{1}{2}}, \tag{68}$$

where p is the Markov model order, and $\widehat{\mathrm{E}}(.;\widehat{\boldsymbol{\vartheta}})$ is the conditional expectation with the parameter vector equal to the MLE of the model.

Note that the conditional expectation $\widehat{\mathrm{E}}(.;\widehat{\boldsymbol{\vartheta}})$ in (68) is an estimation of $\mathrm{E}(.;\boldsymbol{\vartheta})$, which can be used to predict the value of Y_t for $t = 2, \ldots, n$. Also, this expectation depends on the choices of the marginal distributions and the copula family. A general form of the conditional expectation is given by

$$
\begin{aligned}
&\mathrm{E}(Y_t|Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p}; \boldsymbol{\vartheta}) = \\
&\sum_{y_t \in \mathcal{S}} y_t \mathrm{Pr}(Y_t = y_t | Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p}; \boldsymbol{\vartheta}) = \\
&\sum_{y_t \in \mathcal{S}} y_t \frac{\mathrm{Pr}(Y_t = y_t, Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p}; \boldsymbol{\vartheta})}{\mathrm{Pr}(Y_{t-1} = y_{t-1}, \ldots, Y_{t-p} = y_{t-p}; \boldsymbol{\vartheta})}.
\end{aligned}
$$

Residuals analysis is also an alternative choice to evaluate the model fit. However, for dependent discrete data, standard methods, the difference between the predicted and the fitted values, provide residuals that depart from the normal distribution. Dunn and Smyth (1996) introduced randomized quantile residuals that identically and independently distributed as standard normal even if the responses are discrete and dependent. These residuals are given in (48).

4.5 SIMULATED EXAMPLES

To evaluate the performance of the proposed method and confirm the asymptotic results, a comprehensive simulation study was conducted. We carry out the simulation in the statistical software R (R Core Team, 2013). Out of the several processes to choose from, we simulate first order stationary Markov processes with joint distribution of consecutive observations following the bivariate Gaussian and Frank copulas. The marginal distributions are chosen to be the ZIP, ZINB and ZICMP distributions. Since we assume the process is stationary, we set the marginal distributions' parameters, $\boldsymbol{\theta}$, to be constant across time. For Gaussian copula, the marginal parameters are chosen to be

- ZIP with $\boldsymbol{\theta} = (\lambda = 3, \omega = 0.3)'$;

- ZINB with $\boldsymbol{\theta} = (\lambda = 3, \omega = 0.3, \kappa = 5)'$;
- ZICMP with $\boldsymbol{\theta} = (\lambda = 3, \omega = 0.2, \kappa = 0.5)'$;

and the dependence parameter for the bivariate Gaussian copula is chosen to be $\delta = 0.5$ across all three models. We generate 500 simulated datasets for each of the above models with the sample sizes, $n = 100, 200$ and 500. The evaluation criterion is chosen to be the mean absolute deviation error (MADE), which is given by:

$$\frac{1}{m} \sum_{i=1}^{m} |\widehat{\vartheta}_i - \vartheta|,$$

where m is the number of replications, i.e. $m = 500$.

The parameter estimates were obtained after constructing the log-likelihood function given in (61) for the ZIP, ZINB, and ZICMP distributions. A summary of the simulation results are shown in Table 7, which represents the count time series ZIP, ZINB, and ZICMP models with joint distribution of consecutive observations following the bivariate Gaussian copula. The results indicate that the proposed estimation method produces reasonable estimates and relatively small MADEs. In addition, as the sample size increases the parameter estimates seem to converge to the true parameter values. The box plots displayed in figures 26, 27, and 28 show how the performance enhances when the sample size increases. To assess the approximate normality of the estimates, Q-Q plots of the ML estimates for the 500 ZIP, ZINB and ZICMP replicates of length $n = 500$ are shown in figures 29, 30, and 31. These plots agrees with the asymptotic results given in Section 4.3.3.

Table 7: Mean of estimates, MADEs (within parentheses) for Markov zero-inflated models with Gaussian copula

Model	n	λ	ω	κ	δ
ZIP	100	2.9992(0.2773)	0.2949(0.0591)		0.4840(0.0748)
	200	3.0056(0.1698)	0.2961(0.0402)		0.4903(0.0542)
	500	2.9980(0.1051)	0.2977(0.0249)		0.4933(0.0342)
ZINB	100	3.0078(0.3225)	0.2961(0.0661)	4.9338(1.6244)	0.4809(0.0801)
	200	3.0109(0.2252)	0.2968(0.0581)	5.2743(1.4477)	0.4860(0.0581)
	500	2.9990(0.1430)	0.2980(0.0290)	4.9858(1.0384)	0.4913(0.0362)
ZICMP	100	3.4689(0.6813)	0.2008(0.0488)	0.5516(0.0860)	0.4771(0.0747)
	200	3.3383(0.4619)	0.2016(0.0332)	0.5404(0.0611)	0.4847(0.0540)
	500	3.2545(0.2996)	0.2031(0.0210)	0.5326(0.0404)	0.4885(0.0336)

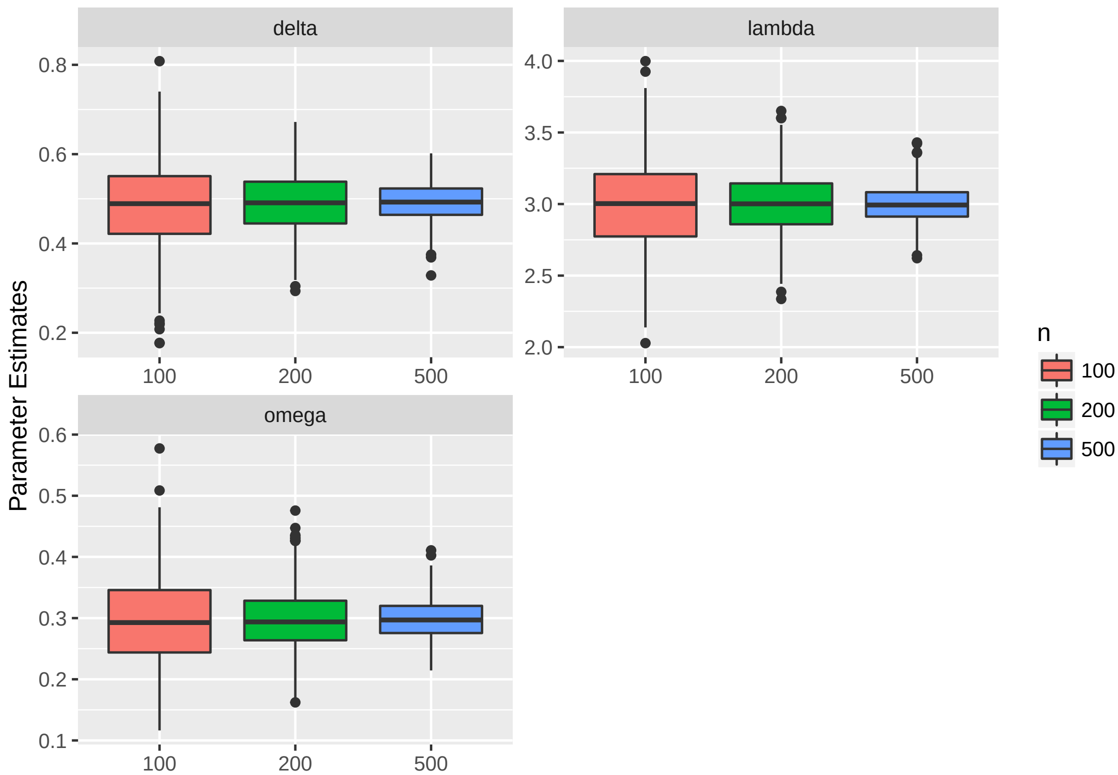


Figure 26: Mean of ML estimates for the ZIP-Gaussian models of length $n = 100, 200,$ and 500

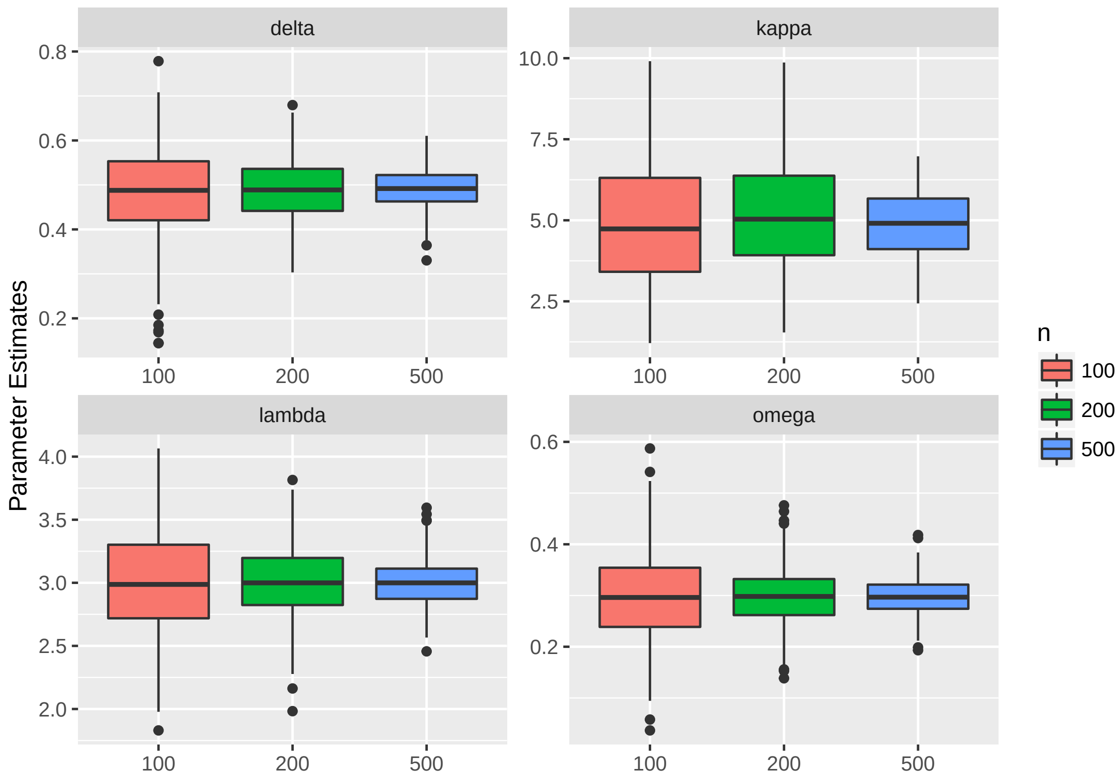


Figure 27: Mean of ML estimates for the ZINB-Gaussian models of length $n = 100, 200,$ and 500

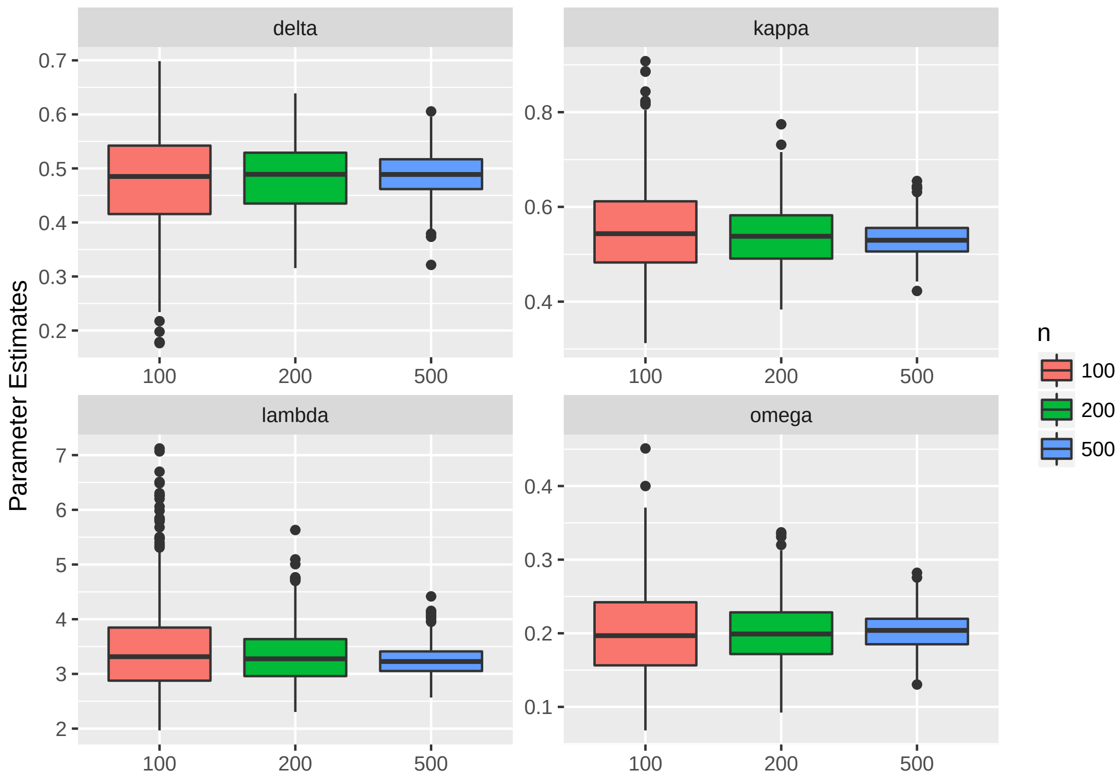


Figure 28: Mean of ML estimates for the ZICMP-Gaussian models of length $n = 100, 200,$ and 500

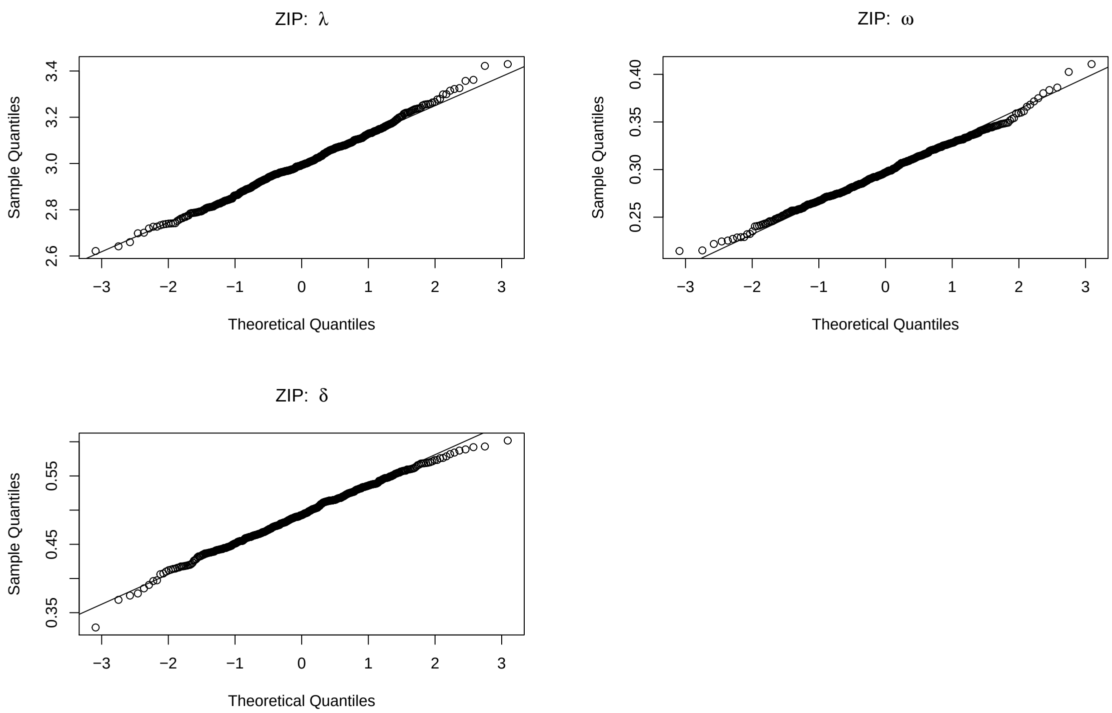


Figure 29: Q-Q plots of the ML estimates for the 500 ZIP-Gaussian process of length $n = 500$.

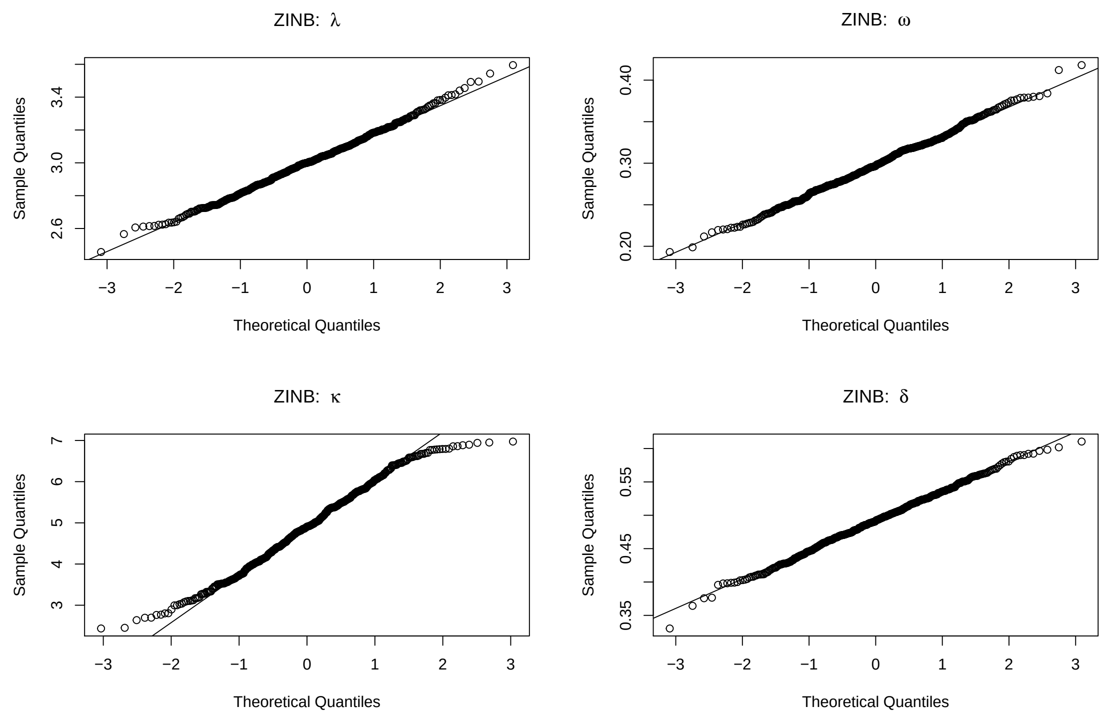


Figure 30: Q-Q plots of the ML estimates for the 500 ZINB-Gaussian process of length $n = 500$.

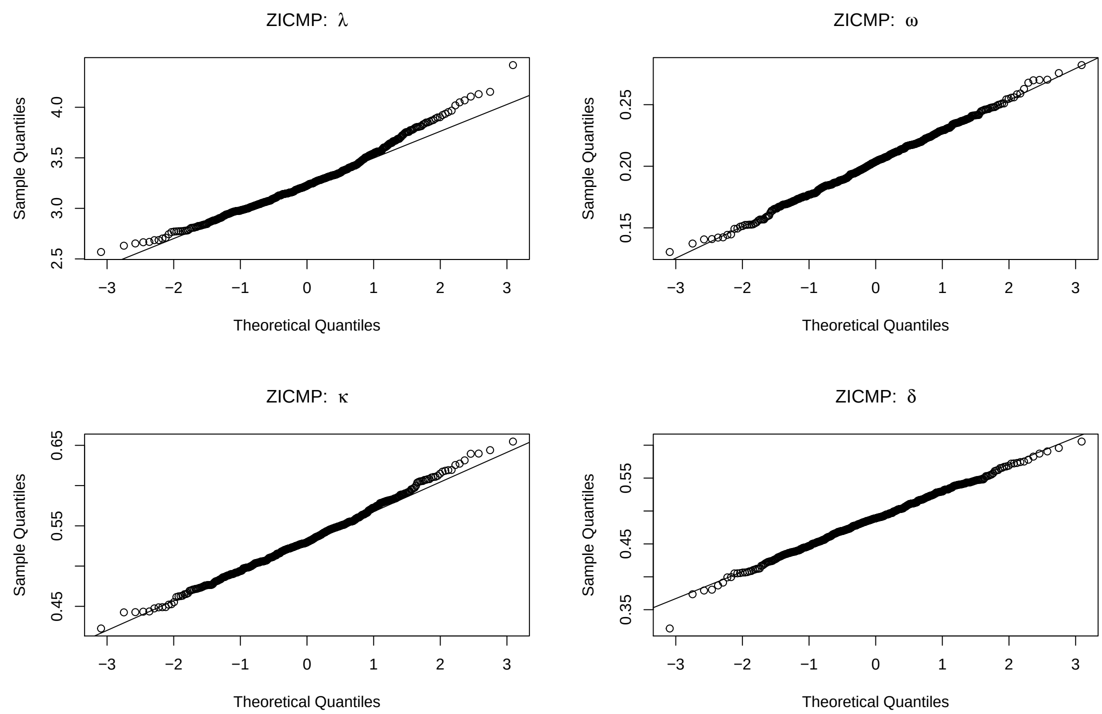


Figure 31: Q-Q plots of the ML estimates for the 500 ZICMP-Gaussian process of length $n = 500$.

Next, we simulate a stationary Markov process with dependence structure following the Frank copula. The marginal and dependence parameters are chosen to be

- ZIP with $\boldsymbol{\theta} = (\lambda = 3, \omega = 0.3)'$;
- ZINB with $\boldsymbol{\theta} = (\lambda = 4.1, \omega = 0.25, \kappa = 0.5)'$;
- ZICMP with $\boldsymbol{\theta} = (\lambda = 4.1, \omega = 0.25, \kappa = 0.5)'$;

and the dependence parameter for the bivariate Frank copula is chosen to be $\delta = 3.2$ across all three models. A summary of the simulation results are shown in Table 8, which represents the count time series ZIP, ZINB, and ZICMP models with joint distribution of consecutive observations following the bivariate Frank copula. Similar to the Gaussian copula, the proposed method performs well. The box plots displayed in figures 32, 33, and 34 indicate that the increasing in sample size improves the estimation. Figures 35, 36, and 37 show the Q-Q plots of the estimates with sample size $n = 500$. The normality assumption of the parameter estimates is satisfied most of the time.

Table 8: Mean of estimates, MADEs (within parentheses) for Markov zero-inflated models with Frank copula

Model	n	λ	ω	κ	δ
ZIP	100	2.9897(0.2513)	0.2935(0.0556)		3.1585(0.5723)
	200	3.0272(0.1916)	0.2902(0.0388)		3.1809(0.5316)
	500	3.0409(0.1051)	0.2965(0.0249)		3.2226(0.3563)
ZINB	100	4.1530(1.1330)	0.2555(0.1434)	0.6780(0.2990)	3.0515(0.7628)
	200	4.0739(0.8759)	0.2471(0.1177)	0.5307(0.1157)	3.1333(0.6089)
	500	4.1200(0.5459)	0.2528(0.0768)	4.9858(1.0384)	3.1394(0.3645)
ZICMP	100	4.3895(0.5621)	0.2493(0.0509)	0.5167(0.0719)	3.2445(0.5745)
	200	4.1753(0.4619)	0.2500(0.0355)	0.5082(0.0490)	3.2312(0.4222)
	500	4.1159(0.3890)	0.2486(0.0203)	0.5052(0.0339)	3.2431(0.2785)

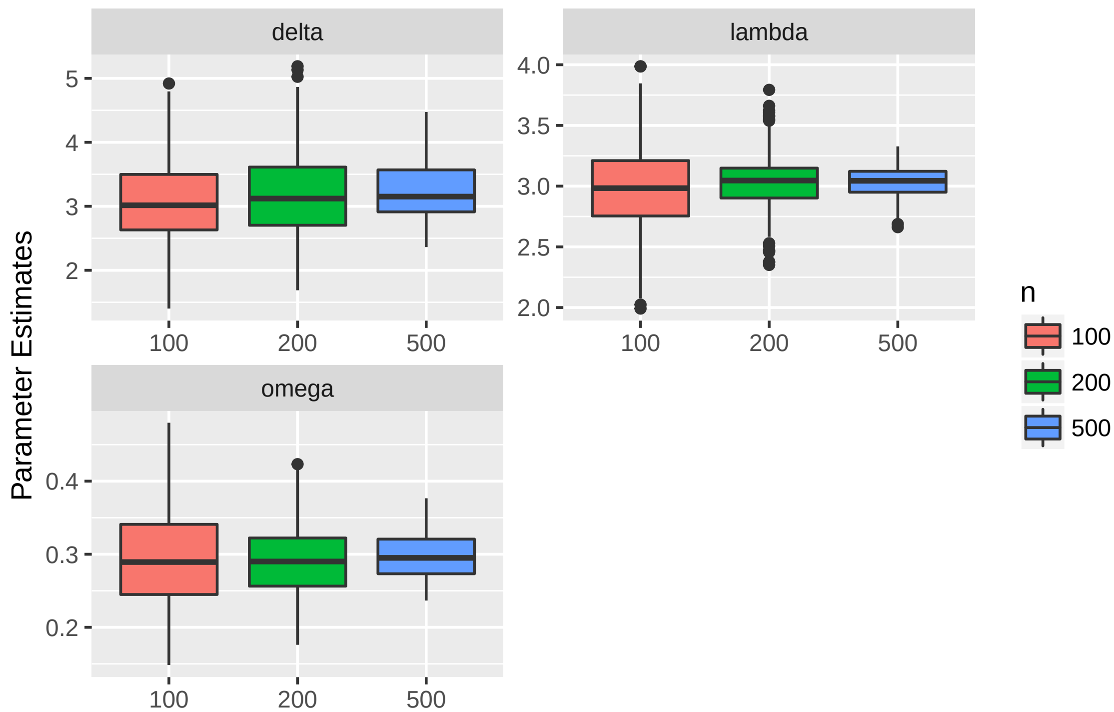


Figure 32: Mean of ML estimates for the ZIP-Frank models of length $n = 100, 200,$ and 500

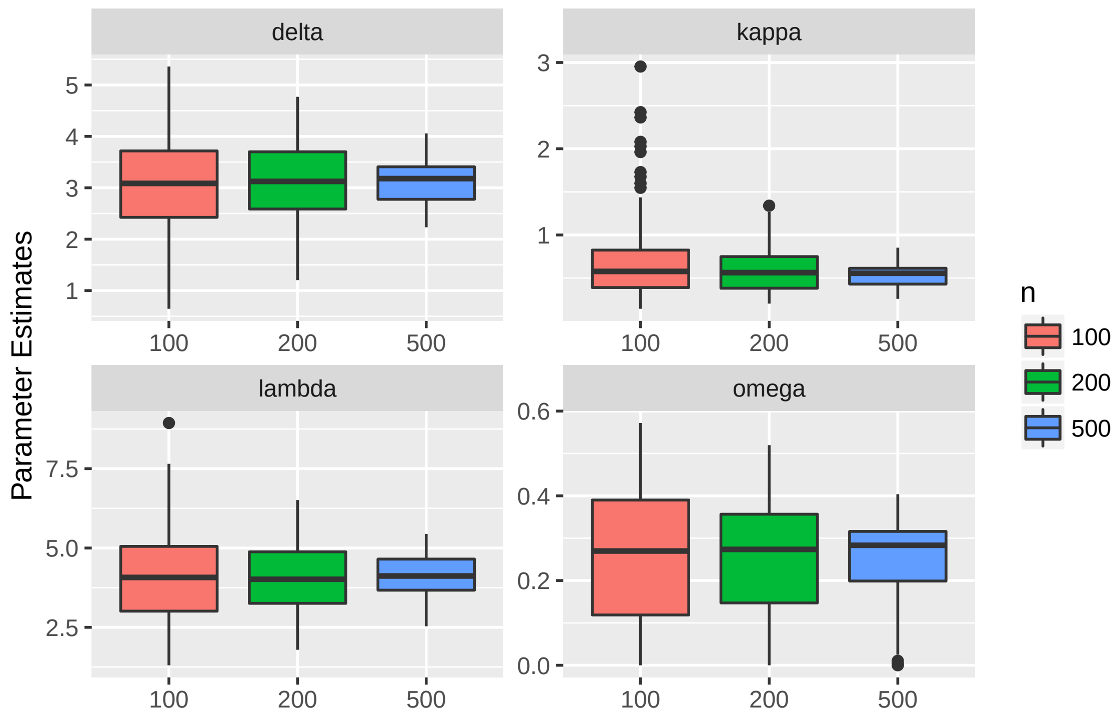


Figure 33: Mean of ML estimates for the ZINB-Frank models of length $n = 100, 200,$ and 500

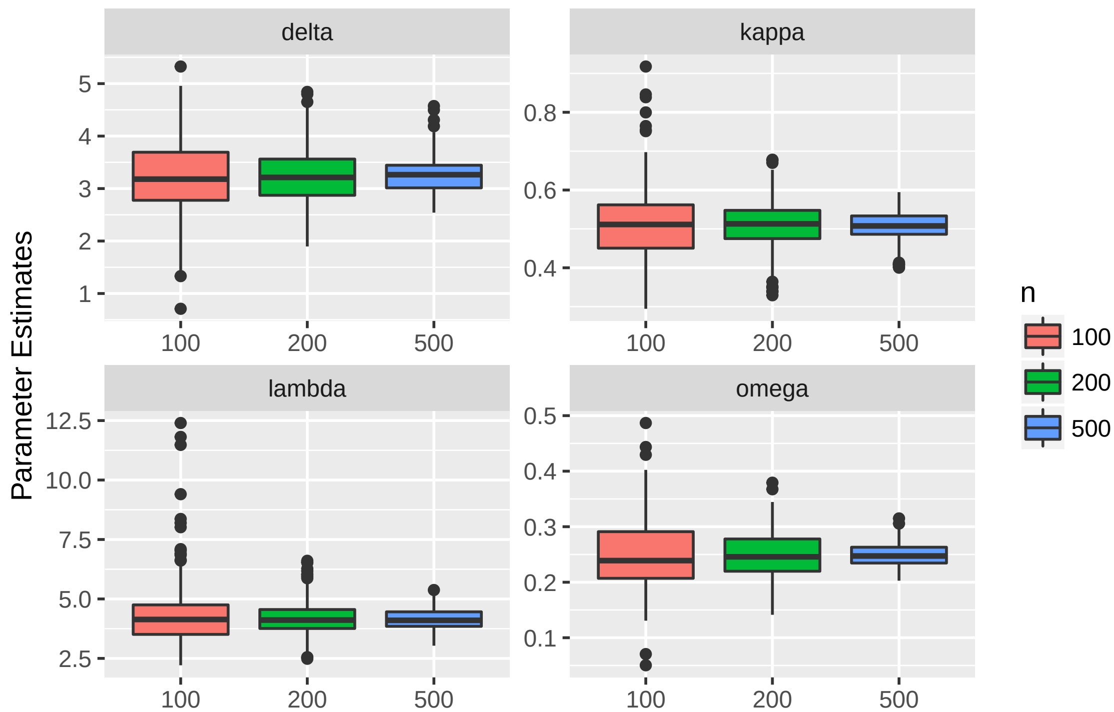


Figure 34: Mean of ML estimates for the ZICMP-Frank models of length $n = 100, 200,$ and 500

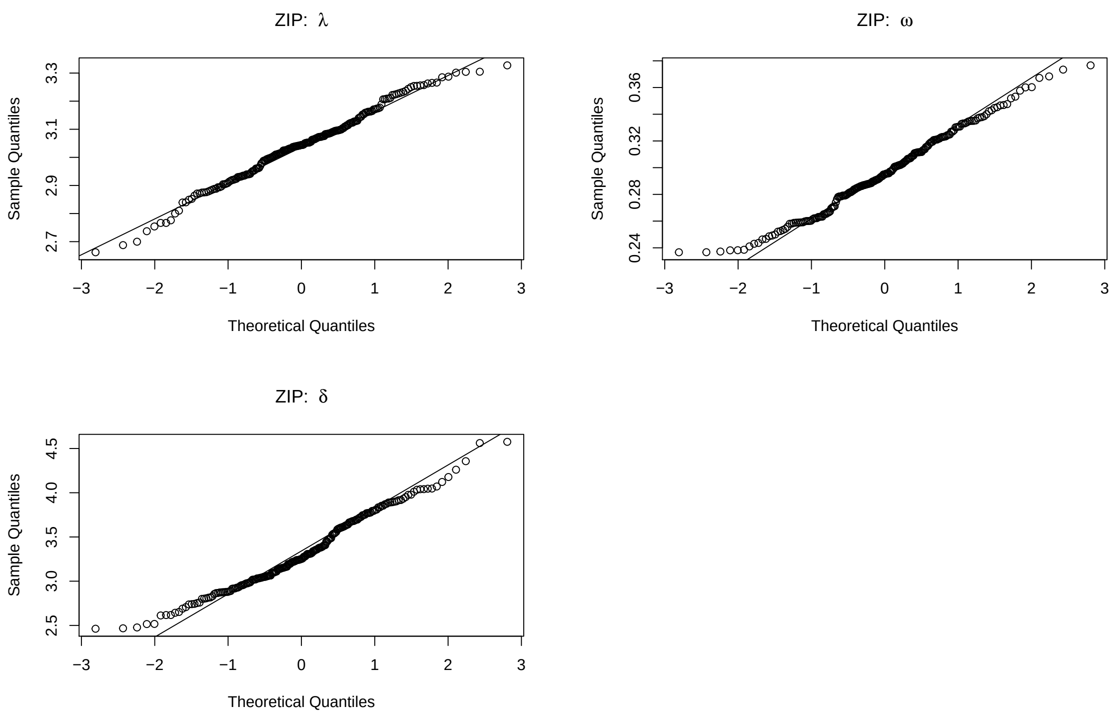


Figure 35: Q-Q plots of the ML estimates for the 500 ZIP-Frank process of length $n = 500$.

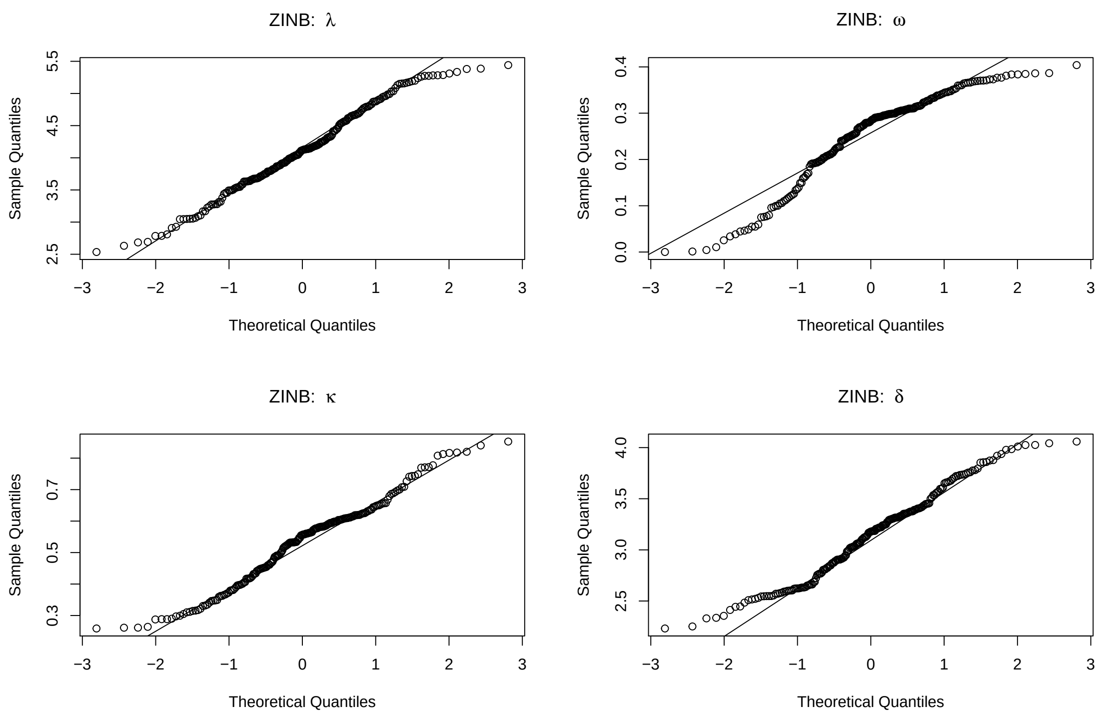


Figure 36: Q-Q plots of the ML estimates for the 500 ZINB-Frank process of length $n = 500$.

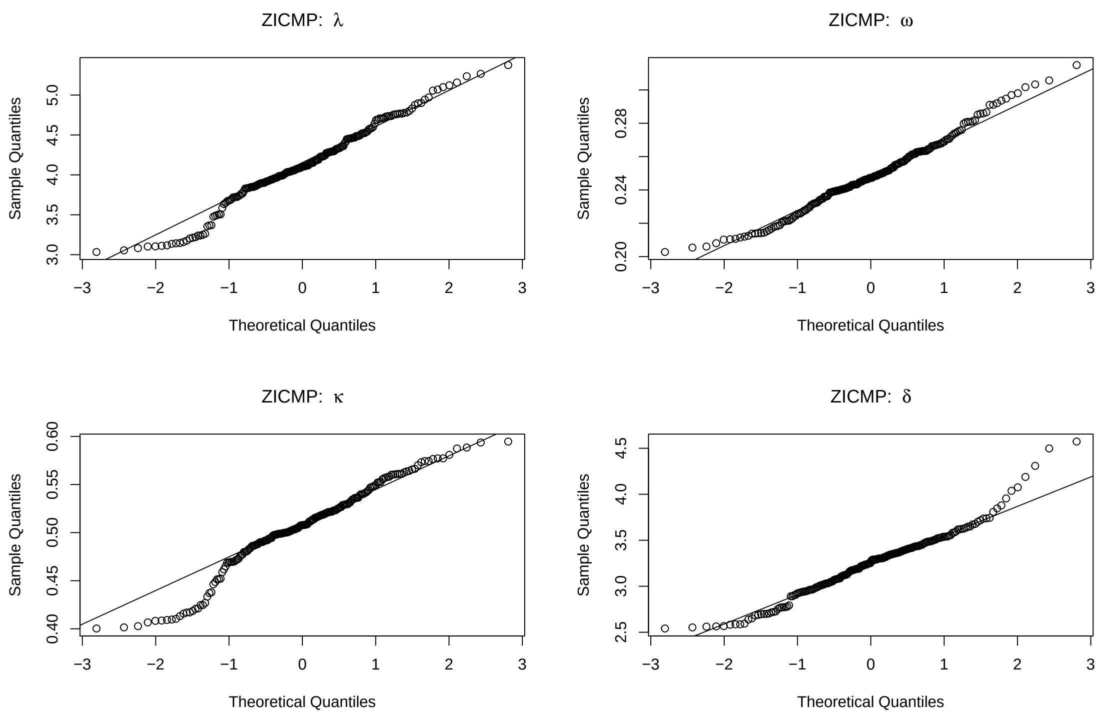


Figure 37: Q-Q plots of the ML estimates for the 500 ZICMP-Frank process of length $n = 500$.

4.6 APPLICATIONS

In this section, we discuss possible applications of the zero-inflated Markov models presented in this chapter. First, we consider monthly counts of arson in the 23rd police car beat plus in Pittsburgh, PA. The data is similar to the one studied in Zhu (2012), and it is from the Forecasting Principles site (http://www.forecastingprinciples.com). Second, we consider the sandstorm data discussed in Chapter 3. The data set consists of the monthly count of strong sandstorms recorded by the AQI airport station in Eastern Province, Saudi Arabia.

4.6.1 ARSON DATA

The data were monthly counts of arson in the 23rd police car beat plus in Pittsburgh, PA. The data consisted of 144 monthly counts of arsons, starting from January 1990 and ending in December 2001. Empirical mean and variance of these zero inflated counts were 1.1042 and 1.5625, respectively. Additionally, empirical mean and variance of these counts, excluding zeros, were 1.8706 and 1.2092, respectively, which suggested non-zero counts observed under-dispersion. A bar plot of the distribution of series is displayed in Figure 38, from which we can see that the distribution of the time series of the arson counts has more zeros relative to a Poisson distribution with the same empirical mean. These zeros represented about 41% of the sample. The count series and the sample autocorrelation function of the series are shown in Figure 39. The plots show that there exist frequent occurrence of zeros and low ordered autocorrelation.

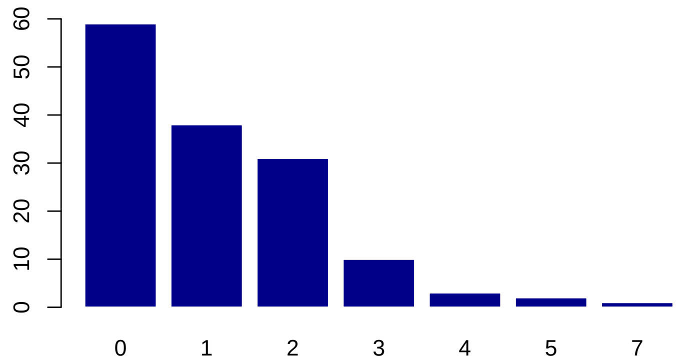


Figure 38: Bar plot of the arson counts series

Hence, we fitted several models with different marginal distributions and dependence structures. The marginal distributions were chosen to be the ZIP, ZINB and ZICMP distributions with constant marginal parameters, i.e. no covariates were considered here. We considered both first order and second order dependence structures. In the first order Markov models, we fitted the bivariate Gaussian, Frank, Gumbel, reflected Gumbel, and Plackett copula functions for the joint distribution of two consecutive observations. For the second order Markov models, we fitted the trivariate Gaussian and max-id copula functions for the joint distribution of three consecutive observations. In the trivariate max-id copula function, we chose the Laplace transform function, $\psi(.)$, to be either the positive staple Laplace transform (PSLT) or the log series Laplace transform (LSLT) with $H(.;\boldsymbol{\delta})$ chosen to be the bivariate Frank copula. Out of these models, we selected three models, each with different marginals, based on the model selection criteria AIC_2, BIC, and RMSPE defined in Section 4.4. Additionally, we considered fitting the ordinary Poisson and negative binomial (NB) distribution in the seek of comparison.

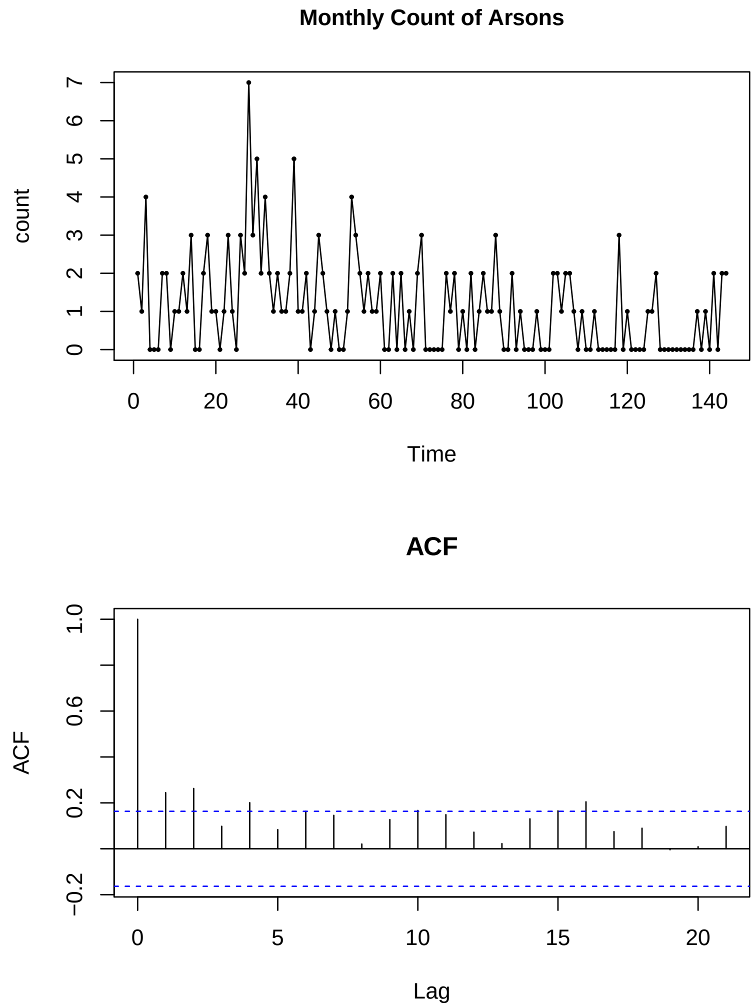


Figure 39: Arson counts series: the time plot, the sample autocorrelation and partial autocorrelation function.

Table 9 shows comparisons of the ZIP, ZINB, ZICMP, Poisson, and NB Markov models with the different dependence structures. The increase of the dependence order improved the AIC and BIC. That is, the second order Markov models outperformed the first order models. Also, accounting for the zero-inflation led to better AIC and BIC values except for the ZICMP. However, in terms of the RMSPE, the ZICMP Markov models were superior to any alternatives, and that might be due to the under-dispersion among the non-zero counts observed in the data. Within each marginal, the values of the RMSPE were not significantly different, but worth noticing that for the ZICMP the first order models outperformed the second order models.

Table 9: Comparisons of the ZIP, ZINB, ZICMP, Poisson, and NB models with different dependence structures

Marginal	Copula	Order	AIC_2	BIC	RMSPE
ZIP	Gaussian	1	198.46	-388.02	1.05
	Frank	1	199.39	-389.86	1.05
	Gumbel	1	201.31	-393.71	1.07
	ref.Gumbel	1	199.30	-389.69	1.06
	Plackett	1	199.56	-390.20	1.05
	Gaussian	2	194.35	-376.83	1.03
	PSLT/Frank	2	195.59	-379.29	1.02
	LSLT/Frank	2	195.39	-378.91	1.03
ZINB	Gaussian	1	197.11	-382.36	1.05
	Frank	1	198.03	-384.18	1.05
	Gumbel	1	199.21	-386.54	1.05
	ref.Gumbel	1	197.95	-384.03	1.05
	Plackett	1	198.18	-384.49	1.05
	Gaussian	2	193.16	-371.48	1.03
	PSLT/Frank	2	193.62	-372.39	1.00
	LSLT/Frank	2	194.00	-373.15	1.03
ZICMP	Gaussian	1	211.38	-410.89	1.51
	Frank	1	212.02	-412.16	1.52
	Gumbel	1	215.32	-418.76	1.48
	ref.Gumbel	1	212.02	-412.17	1.51
	Plackett	1	212.23	-412.59	1.51
	Gaussian	2	202.91	-390.97	1.30
	PSLT/Frank	2	205.65	-396.46	1.17
	LSLT/Frank	2	204.67	-394.49	1.30
Poisson	Gaussian	1	203.55	-401.16	1.21
	Gaussian	2	199.47	-390.02	1.39
NB	Gaussian	1	198.65	-388.40	1.21
	Gaussian	2	195.39	-378.89	1.48

Taking account the values of the AIC, BIC, and RMSPE, we selected the second order Markov models with the trivariate Gaussian copula to handle the dependence structure for the ZIP, ZINB, ZICMP, Poisson, and NB marginals. Table 10 shows the parameter estimates

and standard errors of these models. Out of these models, the ZIP and ZINB distributions were the best in terms of the values of the AIC and BIC. On the other hands, the ZICMP was superior to the rest in terms of the RMSPE with a value equaled to 0.72. The corresponding RMSPE values of the ZIP and ZINB distributions were 1.00 and 1.01, respectively, which were less than the values corresponding to the Poisson and NB distributions.

Table 10: ML estimates, standard errors (within parentheses) for the Markov models. Note: $\beta = \log(\lambda), \gamma = \text{logit}(\omega),$ and $\alpha = \log(\kappa)$

Model	β	γ	α	δ_1	δ_2
Poisson	0.1284(0.0084)			0.2186(0.0265)	0.2301(0.0657)
NB	0.1136(0.0230)		2.7152(0.0259)	0.2677(0.0110)	0.2797(0.0027)
ZIP	0.3567(0.0465)	-1.2392(0.0264)		0.2663(0.0653)	0.2896(0.1314)
ZINB	0.2843(0.0119)	-1.5887(0.0169)	2.1001(0.0151)	0.2704(0.0234)	0.2934(0.0112)
ZICMP	1.7793(0.0360)	-0.3658(0.0036)	2.2782(0.0202)	0.2502(0.0088)	0.2853(0.0115)

Figure 40 displays the predicted values, which were the conditional expectations of Y_t given Y_{t-1} and Y_{t-2} for $t = 1, \ldots, n$ following the ZIP, ZINB, and ZICMP marginals. Within the zero-inflation distributions, the ZIP and ZINB produce similar predicted values.

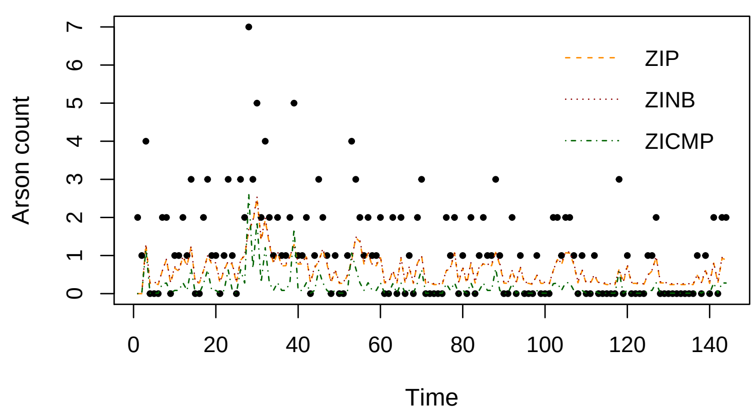


Figure 40: Predicted values using the conditional expectations of the Markov models fit to the arson count series. Dots represent the observed counts.

4.6.2 SANDSTORMS DATA

The data set discussed in this section consists of the 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 main objective was to apply the proposed models and investigate if there were any significant seasonal and trend components. Additionally, we investigated if there were any other predictors that affected the frequency of sandstorms such as the monthly counts of dust haze events, maximum wind speed, temperature, and relative humidity.

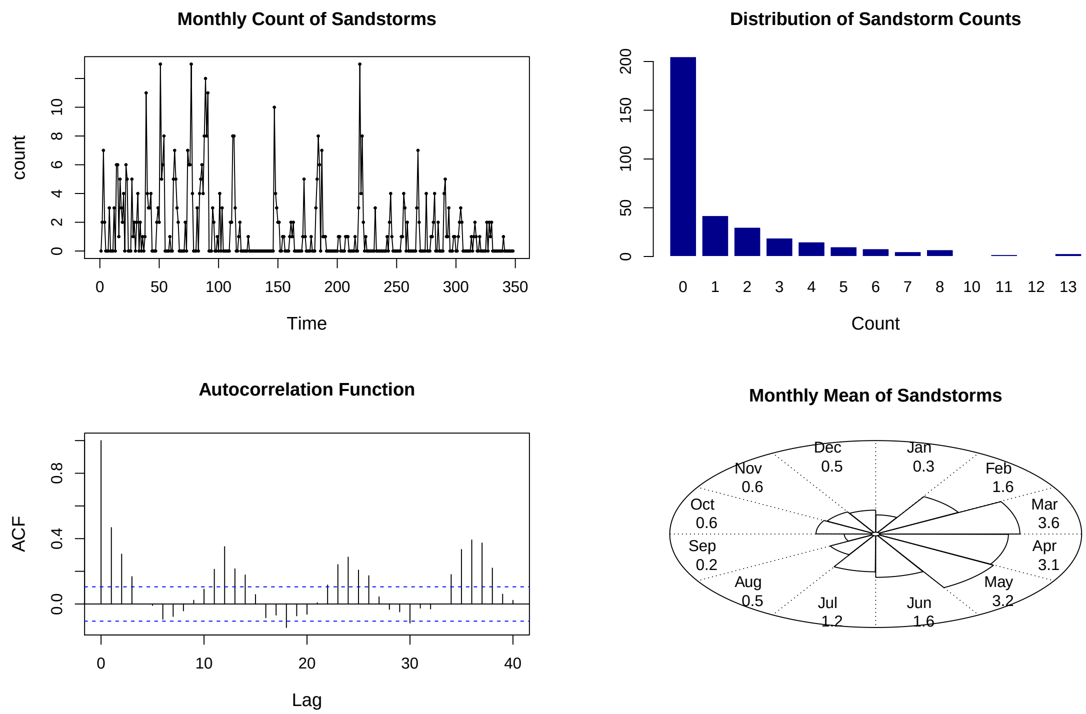


Figure 41: Time series plot of monthly count of sandstorms, the autocorrelation function, bar-plot of distribution of sandstorm counts, and circular plot of the monthly mean count of sandstorms.

Figure 41 shows the sandstorms series plot, the autocorrelation function, bar-plot of the distribution of sandstorm counts, and circular plot of the monthly mean count of sandstorms. From the time series plot and the bar-plot, we could see that the distribution of the sandstorm counts had more zeros relative to a Poisson distribution with the same empirical mean. These zeros represented about 59% of the sample. Decreasing trend could also be observed from the time series plot. Additionally, seasonality was also seen from the autocorrelation function and circular plot. In fact, from the circular plot, we concluded that most sandstorms

occurred during spring time, i.e. March, April, and May months. Thus, trend and seasonal covariates were added to the models.

Hence, we fitted several models with different marginal distributions and dependence structures. The marginal distributions were chosen to be the ZIP, ZINB and ZICMP distributions with the log-linear function of the intensity parameter given by

$$\log(\lambda_t) = \beta_0 + \beta_1 \left(t \times 10^{-3}\right) + \beta_2 x_{1t} + \beta_3 x_{2t} + \beta_4 x_{3t},$$

and the logit function for the zero-inflation parameter given by

$$\text{logit}(\omega_t) = \gamma_0 + \gamma_1 z_{1t} + \gamma_2 z_{2t} + \gamma_3 z_{3t},$$

for $t = 1, \ldots, n$, where $x_{1t} = z_{1t} = \cos\left(\frac{2\pi t}{12}\right)$, $x_{2t} = z_{2t} = \sin\left(\frac{2\pi t}{12}\right)$, and $x_{3t} = z_{3t}$ is the monthly count of dust haze events. The log-function of the dispersion parameter (if existed) given by $\log(\kappa) = \alpha$, i.e. it was chosen to be constant across time. We considered both first order and second order dependence structures. In the first order Markov models, we fitted the bivariate Gaussian, Frank, Gumbel, reflected Gumbel, and Plackett copula functions for the joint distribution of two consecutive observations. For the second order Markov models, we fitted the trivariate Gaussian and max-id copula functions for the joint distribution of three consecutive observations. In the trivariate max-id copula function, we chose the Laplace transform function, $\psi(.)$, to be either the positive staple Laplace transform (PSLT) or the log series Laplace transform (LSLT) with $H(.;\boldsymbol{\delta})$ chosen to be either the bivariate Frank or Gumbel copulas. Out of these models, we selected two models, each with different marginals, based on the model selection criteria AIC$_2$, BIC, and RMSPE defined in Section 4.4.

Table 11 shows comparisons of the ZIP, ZINB, and ZICMP Markov models with the different dependence structures. The increase of the dependence order improves the models. The second order Markov models outperformed the first order Markov models in term of the AIC$_2$, BIC and RMSPE values. However, the second order parameters of the trivariate Gaussian copula and the Laplace transform parameters, in the trivariate max-id copula, are not always significant and dropped if necessary. Additionally, having the same dependence structure, the Markov models with ZINB margins seem to fit the sandstorm data better than the models with ZIP and ZICMP margins. Within the ZIP and ZINB margins, the bivariate reflected Gumbel and Frank copula function are chosen to model the dependence

structures. The trivariate Gaussian copula and the trivariate max-id copula with PSLT and bivariate Gumbel copula function for the ZICMP margin. Hence, we want to show how each dependence structures can be interpreted in term of the autocorrelation.

Table 12 shows that the zero-inflated Markov models are capable of accounting for first order dependence. However, only the models with ZICMP margins account for the second order dependence. The autocorrelation coefficients are similar when the dependence structure is the same for the models with ZIP and ZINB margins. For the marginal parameters, $\boldsymbol{\theta}$, the estimates are quite similar between the ZIP and ZINB, and slightly different from the ZICMP. All models suggest significant decreasing trend in the number of strong sandstorms since $\beta_1 < 0$. Seasonality also significant at annual frequencies since $\beta_2, \beta_3, \gamma_1$ and γ_2 are significantly different from zero. Finally, the affect of dust haze is significant since both β_4 and γ_3 are significantly different from zero. To compare between the Markov models in term of the dependence, we consider the Kendall's tau. That is, the Kendall's tau, when the chosen copula function is the reflected Gumbel, is given by

$$\tau_K = 1 - \delta_1^{-1},$$

so for the ZIP margin it equals to $\tau_K = 0.1908$, which is similar to the one corresponding to the ZINB margin, $\tau_K = 0.2156$. When the copula function is the Frank, the Kendall's tau is then given by

$$\tau_K = 1 - 4\frac{1 - D_1(\delta_1)}{\delta_1},$$

where $D_1(.)$ is the Debye function

$$D_1(x) = \frac{1}{x}\int_0^x \frac{t}{e^t - 1}dt.$$

Thus, for the ZIP margin it equals to $\tau_K = 0.1664$, which is similar to the one corresponding to the ZINB margin, $\tau_K = 0.1870$. In both cases, the ZINB distribution provides slightly stronger dependence than the ZIP distribution.

Table 11: Comparisons of the ZIP, ZINB, and ZICMP models with different dependence structures

Marginal	Copula	Order	AIC_2	BIC	RMSPE
ZIP	Gaussian	1	435.55	-828.72	1.59
	Frank	1	434.11	-825.84	1.58
	Gumbel	1	438.32	-834.28	1.60
	ref.Gumbel	1	433.87	-825.37	1.57
	Plackett	1	435.21	-828.05	1.58
	Gaussian	2	433.85	-821.48	1.56
	PSLT/Frank	2	431.25	-820.13	1.56
	LSLT/Frank	2	433.77	-825.16	1.55
	PSLT/Gumbel	2	418.46	-794.54	1.62
ZINB	Gaussian	1	425.86	-809.34	1.59
	Frank	1	424.54	-806.71	1.59
	Gumbel	1	428.20	-814.03	1.60
	ref.Gumbel	1	424.15	-805.93	1.58
	Plackett	1	425.5	-808.64	1.59
	Gaussian	2	424.06	-801.89	1.56
	PSLT/Frank	2	417.00	-787.78	1.57
	LSLT/Frank	2	423.67	-801.11	1.57
	PSLT/Gumbel	2	413.05	-779.87	1.63
ZICMP	Gaussian	1	477.10	-911.82	1.91
	Frank	1	475.06	-907.75	1.88
	Gumbel	1	473.49	-904.6	1.79
	ref.Gumbel	1	473.73	-905.1	1.87
	Plackett	1	469.06	-895.74	1.77
	Gaussian	2	449.97	-853.71	1.72
	PSLT/Frank	2	457.44	-868.66	1.80
	LSLT/Frank	2	450.91	-855.59	1.77
	PSLT/Gumbel	2	455.12	-864.02	1.83

Table 12: Parameter estimates (standard errors) for the copula-based Markov models fit to the sandstorms count series.

	ZIP		ZINB		ZICMP	
Parameter	ref.Gumbel	Frank	ref.Gumbel	Frank	Gaussian	PSLT/Gumbel
β_0	0.9578(0.1190)	0.9705(0.1172)	0.9960(0.1564)	0.9391(0.1593)	0.8139(0.0577)	0.7701(0.1938)
β_1	-4.2579(0.6098)	-4.0715(0.6049)	-4.9031(0.8052)	-4.7335(0.8168)	-2.3656(0.0302)	-2.4943(0.6999)
β_2	-0.2184(0.0890)	-0.1789(0.0879)	-0.1956(0.1233)	-0.1836(0.1253)	-0.0921(0.0107)	-0.1248(0.0860)
β_3	0.3722(0.0957)	0.3650(0.0950)	0.4371(0.1255)	0.4379(0.1283)	0.2203(0.0554)	0.2466(0.0827)
β_4	0.0656(0.0088)	0.0638(0.0089)	0.0635(0.0121)	0.0636(0.0124)	0.0452(0.0024)	0.0426(0.0112)
γ_0	0.6627(0.2717)	0.7264(0.2677)	0.5357(0.3077)	0.5926(0.3086)	1.1825(0.0481)	1.1292(0.2221)
γ_1	0.6236(0.2561)	0.6615(0.2514)	0.6706(0.2991)	0.6607(0.3006)	0.6252(0.0717)	0.6266(0.2014)
γ_2	-0.9051(0.2507)	-0.8798(0.2427)	-0.8819(0.2814)	-0.7799(0.2799)	-0.9827(0.0456)	-0.9824(0.2070)
γ_3	-0.1407(0.0444)	-0.1498(0.0448)	-0.1596(0.0538)	-0.1664(0.0545)	-0.1565(0.0181)	-0.1467(0.0335)
α			1.4876(0.5709)	1.4303(0.5431)	0.8083(0.0186)	0.7540(0.1400)
δ_1	1.2358(0.0765)	1.5326(0.4672)	1.2748(0.0896)	1.7328(0.5034)	0.2779(0.0444)	1.0242(0.0885)
δ_2					0.1818(0.0238)	1.1922(0.1057)
τ_K	0.1908	0.1664	0.2156	0.1870	0.1793	0.1612

Figure 42 displays the predicted values, which were the conditional expectations of Y_t given Y_{t-1} for the ZIP and ZINB Markov models and Y_t given Y_{t-1} and Y_{t-2} for ZICMP Markov models for $t = 1, \ldots, n$. The ZIP and ZINB models perform better than the ZICMP especially with the first hundred observations where non-zero counts are more frequent. Within each margin, the reflected Gumbel and Frank copulas are very similar with the ZIP and ZINB margins. However, with the ZICMP margin, the Gaussian copula is better than the max-id copula.

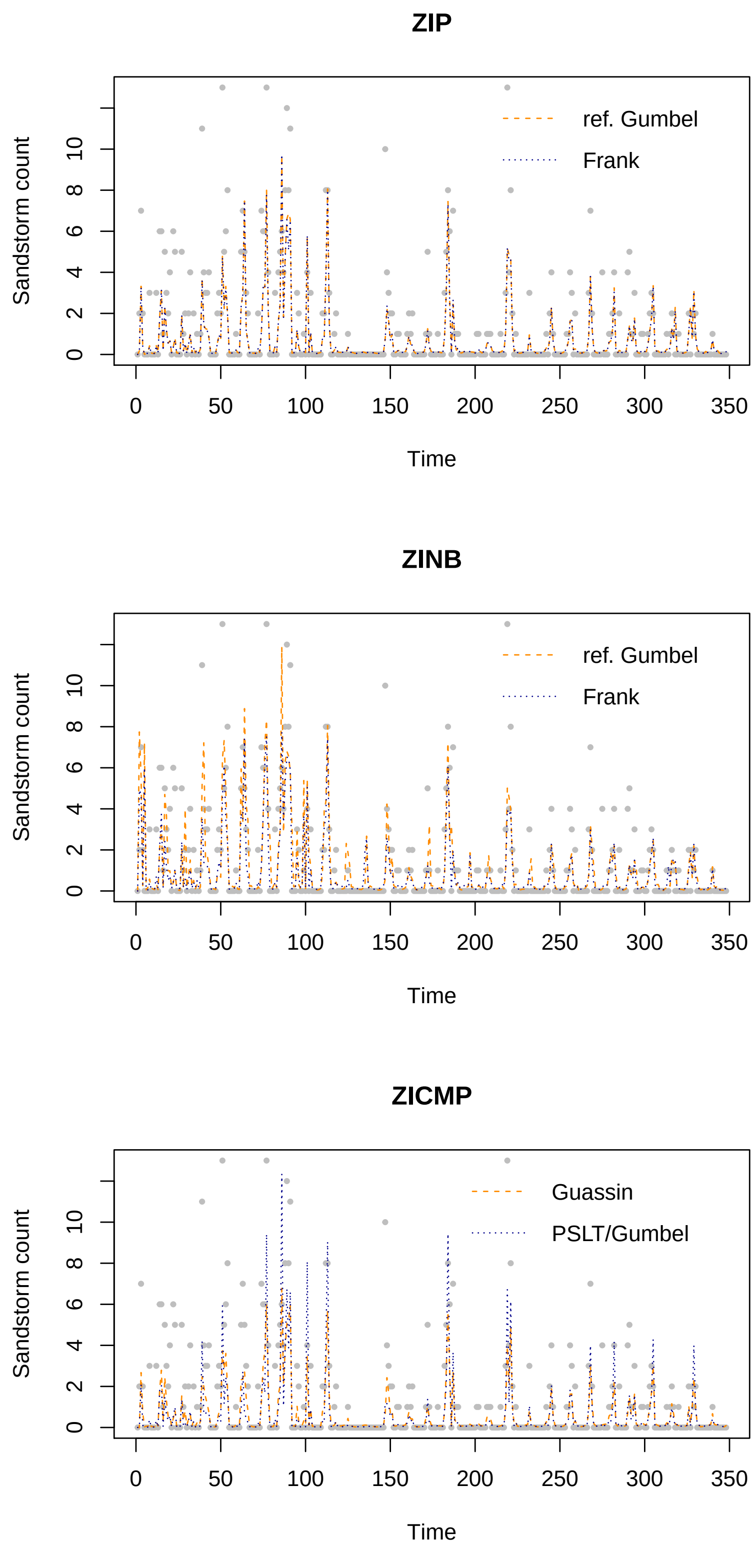


Figure 42: Predicted values using the conditional expectations of the Markov models fit to the sandstorm count series. Dots represent the observed counts.

CHAPTER 5

SUMMARY AND FUTURE DIRECTIONS

5.1 SUMMARY

Count time series data are observed in several applied disciplines such as in environmental science, biostatistics, economics, public health, and finance. In some cases, a specific count, say zero, may occur more often than usual. However, during certain periods these counts could take larger values. Additionally, in practice these zero-inflated counts usually observe serial dependence when the data is collected over time. Overlooking the frequent occurrence of zeros and the serial dependence could lead to false inference.

In this dissertation, we have proposed two classes of copula-based zero-inflated count time series models. The first class used the marginal ZIP, ZINB, and ZICMP regression models for zero-inflated count time series data with the serial dependence being captured by a Gaussian copula, with a correlation matrix corresponding to a stationary autoregressive moving average (ARMA) process. Likelihood inference is carried out using sequential importance sampling. Simulated studies were conducted to evaluate the estimation method. The studies show that the estimation method is accurate and reliable even for relatively small sample size for the ZIP and ZINB models. However, for the ZICMP models, the method is less consistent for smaller sample size. Model assessment to check the goodness of the proposed models is done via residual analysis. The proposed models are applied on the occupational health and sandstorms data. According to the residual analysis the models fit the data adequately. A significant advantage of this model is that interpretations of the model components are easily derived.

The second class also used the marginal ZIP, ZINB, and ZICMP distributions to build Markov zero-inflated count time series models. The serial dependence being captured through constructing bivariate and trivariate joint distributions of the consecutive observations. The joint distribution function of the consecutive observations is constructed through copula functions such as the Gaussian, Frank, Gumbel copula functions. Model properties

and dependence measurements are discussed. We also implemented model selection and prediction to assess the models performance. Simulation studies were conducted to evaluate the estimation method. The studies showed that the estimated parameters are consistent and normally distributed. The proposed Markov models are applied on the arson data and again on the sandstorms data. The models prove to be reliable on handling different zero-inflated count time series data.

5.2 FUTURE DIRECTIONS

There are several directions one can take on related future work of this dissertation. The following are some extensions.

5.2.1 MULTIVARIATE ZERO-INFLATED COUNT TIME SERIES MODELS

In practice, many time series components come as a vector instated of a single observation. For example, in our own sandstorms data, we can expand the data by including other counts from different stations in Saudi Arabia. Hence, we can investigate both the temporal autocorrelation within each time series from each station and the spatial correlation between the stations. That is, we have a multivariate process $\{\mathbf{Y}_t\}$ with dimension d for $t = 1, \ldots, n$. Then, besides the serial dependence within each series $\{Y_{ti}\}$ for $i = 1, \ldots, d$, we have spatial dependence (or interdependence in general) between the different series $\{Y_{ti}\}$ and $\{Y_{tj}\}$ for $i \neq j$ and $t = 1, \ldots, n$.
The covariance matrix of such a process is given by

$$\text{Cov}(\mathbf{Y}_{t+h}, \mathbf{Y}_t) = \begin{bmatrix} \text{Cov}(Y_{t+h,1}, Y_{t,1}) & \text{Cov}(Y_{t+h,1}, Y_{t,2}) & \ldots & \text{Cov}(Y_{t+h,1}, Y_{t,d}) \\ \text{Cov}(Y_{t+h,2}, Y_{t,1}) & \text{Cov}(Y_{t+h,2}, Y_{t,2}) & \ldots & \text{Cov}(Y_{t+h,2}, Y_{t,d}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(Y_{t+h,d}, Y_{t,1}) & \text{Cov}(Y_{t+h,d}, Y_{t,2}) & \ldots & \text{Cov}(Y_{t+h,d}, Y_{t,d}) \end{bmatrix}_{nd \times nd}.$$

Copula theory can be applied to construct such models.

5.2.2 DIRECTIONAL DEPENDENCE OF COUNT TIME SERIES

In some cases, the time series counts come in pairs, and researchers are interested in

studying the dynamic relationships of these pairs. Many of the methods applied to study such dynamic relationships are plausible only if the random processes were continuous or the relationships were symmetric. One of these methods is the Granger-causality test introduced by Granger (1969), which has become a standard technique used to study the causal relationship. Kim and Hwang (2017) introduced a new method of investigating the casual relationships of two asymmetric time series processes by deriving measurements of the directional dependence via a Gaussian copula beta regression model with generalized autoregressive conditional heteroscedasticity (GARCH) marginals. Using directional dependence measurements suggested in Sungur (2005) and Alqawba et al. (2019), we plan to study the joint behavior of bivariate count time series using the proposed models in this dissertation and other models in the literature.

APPENDIX A

A.1 CONDITIONAL DISTRIBUTION OF THE LATENT ERROR VECTOR GIVEN THE OBSERVED PROCESS IN CHAPTER 3

In Chapter 3, we used the model

$$Y_t = F_t^{-1}\{\Phi(\epsilon_t)|\mathbf{X}_t;\boldsymbol{\theta}\}, \quad \text{for } t = 1,\dots,n,$$

which is proportional to

$$\Phi^{-1}(F_t\{Y_t - 1|\mathbf{X}_t;\boldsymbol{\theta}\}) < \epsilon_t \leq \Phi^{-1}(F_t\{Y_t|\mathbf{X}_t;\boldsymbol{\theta}\}), \quad \text{for } t = 1,\dots,n,$$

and hence when Y_t takes the value y_t, the latent variable ϵ_t falls in the interval $\mathcal{D}_t(y_t|\mathbf{X}_t;\boldsymbol{\theta}) = (\Phi^{-1}\{F_t(y_t^-|\mathbf{X}_t;\boldsymbol{\theta})\}, \Phi^{-1}\{F_t(y_t|\mathbf{X}_t;\boldsymbol{\theta})\}]$ (Lennon, 2016).

Now, in our model, we assumed that the errors $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, R(\boldsymbol{\rho}))$. Thus, given the data $\mathbf{Y} = \mathbf{y}$, the conditional distribution of the errors is a multivariate truncated normal distribution on the interval $\boldsymbol{D} = (\boldsymbol{D}^-, \boldsymbol{D}^+]$ where

$$\boldsymbol{D}^- = (\Phi^{-1}\{F_1(y_1^-;\boldsymbol{\theta})\},\dots,\Phi^{-1}\{F_n(y_n^-;\boldsymbol{\theta})\})',$$

and

$$\boldsymbol{D}^+ = (\Phi^{-1}\{F_1(y_1;\boldsymbol{\theta})\},\dots,\Phi^{-1}\{F_n(y_n;\boldsymbol{\theta})\})'.$$

The covariate vector $\mathbf{X}_t$ is omitted here for simplicity. The multivariate pdf of $\boldsymbol{\epsilon}$ given $\mathbf{Y} = \mathbf{y}$ is then given by

$$f(\boldsymbol{\epsilon}|\mathbf{Y} = \mathbf{y}) = \begin{cases} \frac{\exp\{-\frac{1}{2}\boldsymbol{\epsilon}' R^{-1}\boldsymbol{\epsilon}\}}{\int_{\boldsymbol{D}^-}^{\boldsymbol{D}^+} \exp\{-\frac{1}{2}\boldsymbol{\epsilon}' R^{-1}\boldsymbol{\epsilon}\} d\boldsymbol{\epsilon}} & \boldsymbol{D}^- < \boldsymbol{\epsilon} \leq \boldsymbol{D}^+ \\ 0 & \text{otherwise,} \end{cases}$$

where $R = R(\boldsymbol{\rho})$ is the positive definite covariance and correlation matrix of the error, i.e. its diagonal values are 1's.
Therefore, the marginal distributions of $\epsilon_t | Y_t = y_t$ for $t = 1, \ldots, n$ are univariate truncated normal on the interval $\mathcal{D}_t = (\mathcal{D}_t^-, \mathcal{D}_t^+]$ with pdf's given by

$$
\begin{aligned}
f(\epsilon_t | Y_t = y_t) &= \begin{cases} \frac{\phi(\epsilon_t)}{\Phi(\mathcal{D}^+) - \Phi(\mathcal{D}^-)} & \mathcal{D}^- < \epsilon_t \leq \mathcal{D}^+ \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} \frac{\phi(\epsilon_t)}{F_t(y_t; \boldsymbol{\theta}) - F_t(y_t^-; \boldsymbol{\theta})} & \mathcal{D}^- < \epsilon_t \leq \mathcal{D}^+ \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}
$$

The expectation of $\epsilon_t | Y_t = y_t$ is then given in (49), which is used to estimate the latent variable ϵ_t.

A.2 TRIVARIATE MAX-ID COPULA FUNCTION WITH POSITIVE STABLE LT AND BIVARIATE GUMBEL

The following is a derivation of the trivariate max-id copula function with positive stable LT and bivariate Gumbel, which results the trivariate extreme value copula function. The positive stable function is given by

$$\psi(s) = \exp\{-s^{1/\delta_1}\}, \quad \delta_1 \geq 1,$$

with corresponding functional inverse given by

$$\psi^{-1}(t) = (-\log t)^{\delta_1}.$$

Hence, the trivariate max-id copula given in (59) becomes

$$
\begin{aligned}
F_{123}(y_t, y_{t-1}, y_{t-2}) &= \exp\Bigg\{ -\Bigg(\sum_{j \in \{t, t-2\}} \Big[\\
&- \log H(e^{-0.5(-\log F_j)^{\delta_1}}, e^{-0.5(-\log F_{t-1})^{\delta_1}}; \delta_2) \\
&+ \frac{1}{2}(-\log F_j)^{\delta_1} \Big] \Bigg)^{1/\delta_1} \Bigg\}. \qquad (69)
\end{aligned}
$$

Now, if $H(.;\delta_2)$ is chosen to be the Gumbel copula, i.e.

$$C(u_1,u_2;\delta)=\exp\{-([-\log u_1]^{\delta}+[-\log u_2]^{\delta})^{1/\delta}\},$$

then (69) becomes

$$\begin{aligned}
F_{123}(y_t,y_{t-1},y_{t-2}) &= \exp\Bigg\{-\Bigg(\sum_{j\in\{t,t-2\}}\Big[-\log\exp\Big\{-\Big([-\log e^{-.5(-\log F_j)^{\delta_1}}]^{\delta_2} \\
&+ [-\log e^{-.5(-\log F_{t-1})^{\delta_1}}]^{\delta_2}\Big)^{1/\delta_2}\Big\}+\frac{1}{2}(-\log F_j)^{\delta_1}\Big]\Bigg)^{1/\delta_1}\Bigg\} \\
&= \exp\Bigg\{-\Bigg(\sum_{j\in\{t,t-2\}}\Big[\Big([\frac{1}{2^{\delta_2}}(-\log F_j)^{\delta_1\delta_2} \\
&+ \frac{1}{2^{\delta_2}}(-\log F_{t-1})^{\delta_1\delta_2}\Big)^{1/\delta_2}+\frac{1}{2}(-\log F_j)^{\delta_1}\Big]\Bigg)^{1/\delta_1}\Bigg\} \\
&= \exp\Bigg\{-\Bigg(\Big[\frac{1}{2^{\delta_2}}(-\log F_t)^{\delta_1\delta_2}+\frac{1}{2^{\delta_2}}(-\log F_{t-1})^{\delta_1\delta_2}\Big]^{1/\delta_2} \\
&+ \Big[\frac{1}{2^{\delta_2}}(-\log F_{t-2})^{\delta_1\delta_2}+\frac{1}{2^{\delta_2}}(-\log F_{t-1})^{\delta_1\delta_2}\Big]^{1/\delta_2} \\
&+ \frac{1}{2}(-\log F_t)^{\delta_1}+\frac{1}{2}(-\log F_{t-2})^{\delta_1}\Bigg)^{1/\delta_1}\Bigg\}. \qquad (70)
\end{aligned}$$

The bivariate margins of (70) is then given by

$$\begin{aligned}
F_{j2}(y_j,y_{t-1}) &= \psi\Bigg(-\log H(e^{-0.5\psi^{-1}(F_j;\delta_1)},e^{-0.5\psi^{-1}(F_{t-1};\delta_1)};\delta_2) \\
&+ \frac{1}{2}\psi^{-1}(F_j;\delta_1)+\frac{1}{2}\psi^{-1}(F_{t-1};\delta_1);\delta_1\Bigg) \\
&= \exp\Bigg\{-\Bigg(\log H(e^{-0.5(-\log F_j)^{\delta_1}},e^{-0.5(-\log F_{t-1})^{\delta_1}};\delta_2) \\
&+ \frac{1}{2}(-\log F_j)^{\delta_1}+\frac{1}{2}(-\log F_{t-1})^{\delta_1}\Bigg)^{1/\delta_1}\Bigg\} \\
&= \exp\Bigg\{-\Bigg(\Big(\frac{1}{2^{\delta_2}}(-\log F_j)^{\delta_1\delta_2}+\frac{1}{2^{\delta_2}}(-\log F_{t-1})^{\delta_1\delta_2}\Big)^{1/\delta_2}
\end{aligned}$$

$$+ \quad \frac{1}{2}(-\log F_j)^{\delta_1} + \frac{1}{2}(-\log F_{t-1})^{\delta_1}\Bigg)^{1/\delta_1}\Bigg\},$$

for $j = t, t-2$, and

$$F_{13}(y_t, y_{t-2}) = \exp\Bigg\{ - \Big[(-\log F_t)^{\delta_1} + (-\log F_{t-2})^{\delta_1}\Big]^{1/\delta_1}\Bigg\}.$$

APPENDIX B

SELECTED R CODES

All computations in this dissertations are written and implemented in the R software (R Core Team, 2013). The following are a selection of R codes used on our work.

B.1 ZIP MARGINAL FUNCTION FOR THE GAUSSIAN COPULA ZERO-INFLATED COUNT TIME SERIES MODEL

The following R function can be used to specify the ZIP marginal distribution when the covariates are present on both the intensity parameter and the zero-inflation parameter as presented in Chapter 3. The function is then can be applied to the R package "gcmr" (Masarotto and Varin, 2017) to obtain the parameter estimates.

```
ZIP.marg.cov <- function(link = "log"){
  fm <- poisson( substitute( link ) )
  ans <- list()
  ans$start <- function(y, x, z, offset) {

    offset <- list( as.vector(offset$mean), as.vector(offset$precision) )
    m <- zeroinfl(y ~ x   | z , EM = F)
    lambda <- coef(m)
    if( is.null(z) ){
      pos <- NCOL(x)+1
      lambda[pos] <- exp( lambda[pos] )
      names(lambda)[pos] <- "Zero"
      attr(lambda, "lower") <- c( rep( -Inf, NCOL(x) ), sqrt(.Machine$double.eps) )
    }
    lambda
  }

  ans$npar <- function(x, z) ifelse(!is.null(z), NCOL(x)+NCOL(z)+2, NCOL(x)+1)
  ans$dp <- function(y, x, z, offset, lambda) {
    nb <- length(lambda)
    mu <- exp( cbind(1,x) %*% lambda[ 1:(NCOL(x)+1) ])
    if( is.null(z) )
      phi <- 0
    else

      phi <- as.numeric((exp( cbind(1,z) %*% lambda[  (NCOL(x)+2):(NCOL(x)+NCOL(z)+2) ] )
          /(1 + exp( cbind(1,z) %*% lambda[  (NCOL(x)+2):(NCOL(x)+NCOL(z)+2) ] ))))
```

```
    cbind(dzipois(y, mu,phi), pzipois(y,mu,(phi)))
  }
  ans$q <- function(p, x, z, offset, lambda) {
    nb <- length(lambda)
    mu <- exp( cbind(1,x) %*% lambda[1:(NCOL(x)+1)] )
    if( is.null(z) )
      phi <- 0
    else
      phi <-as.numeric(exp( cbind(1,z) %*% lambda[  (NCOL(x)+2):(NCOL(x)+NCOL(z)+2) ] )/
          (1 + exp( cbind(1,z) %*% lambda[  (NCOL(x)+2):(NCOL(x)+NCOL(z)+2) ] )))

    qzipois(p, mu, (phi))
  }
  ans$fitted.val <- function(x, z, offset, lambda){
    exp( cbind(1,x) %*% lambda[ 1:(NCOL(x)+1) ] )

  }
  ans$type <- "integer"
  class(ans) <- c("marginal.gcmr")
  ans
}
```

B.2 LOG-LIKELIHOOD FUNCTION OF THE MARKOV ORDER 1 ZIP MODEL

The following R function is the negative log-likelihood of the Markov order 1 ZIP model given in Chapter 4.

```
# Inputs arguments:
# param : parameter vector of the model
# y : a vector of the zero-inflated counts
# X : a matrix of the covariates associated with the intensity parameters
# Z : a matrix of the covariates associated with the zero-inflated parameters
# pcop: bivariate copula cdf
# cparlb: lower bound on the copula parameter
# cparub:  upper bound on the copula parameter
ZIPnllk=function(param,y,X=0,Z=0,pcop=pbvncop,cparlb=0,cparub=30)
{ n=length(y)
X=as.matrix(X)
if(nrow(X)==1) nc1=0 else nc1=ncol(X)
Z=as.matrix(Z)
if(nrow(Z)==1) nc2=0 else nc2=ncol(Z)
#nc1=ncol(X)  # should be same length as bvec (beta vector)
#nc2=ncol(Z)  # should be same length as gvec (gamma vector)

b0=param[1]
```

```
g0=param[(nc1+2)]
np=length(param)
cpar=param[np]
if(cpar<=cparlb | cpar>=cparub) return(1.e10)

if(nc1>0)
{ bvec=param[2:(nc1+1)];
muvec=b0+X%*%bvec; muvec=exp(muvec)
}
else muvec=rep(exp(b0),n)

if(nc2>0)
{ gvec=param[(nc1+3):(nc1+nc2+2)];
omegavec=g0+Z%*%gvec; omegavec=exp(omegavec)/(1+exp(omegavec))
}
else omegavec=rep(exp(g0)/(1+exp(g0)),n)
#thv=(muvec, omegavec)

cdf1=pzipois(y,lambda=muvec,pstr0=omegavec)
pmf=dzipois(y,lambda=muvec,pstr0=omegavec)
cdf0=cdf1-pmf
cdf0[cdf0<=0]=0
nllk=-log(pmf[1])
for(i in 2:n)
{ tem=pcop(cdf1[i-1],cdf1[i],cpar) - pcop(cdf0[i-1],cdf1[i],cpar) -
  pcop(cdf1[i-1],cdf0[i],cpar) + pcop(cdf0[i-1],cdf0[i],cpar)
condpr=tem/pmf[i-1]
if(condpr<=0. | is.na(condpr)) condpr=1.e-15
nllk=nllk-log(condpr)
}
nllk
}
```

B.3 LOG-LIKELIHOOD FUNCTION OF THE MARKOV ORDER 2 ZIP MODEL WITH GAUSSIAN COPULA

The following R function is the negative log-likelihood of the Markov order 2 ZIP model with Gaussian Copula given in Chapter 4.

```
# Inputs arguments:
# param : parameter vector of the model
# y : a vector of the zero-inflated counts
# X : a matrix of the covariates associated with the intensity parameters
# Z : a matrix of the covariates associated with the zero-inflated parameters
GauZIP2nllk=function(param,y,X=0,Z=0)
{
n=length(y)
```

```
X=as.matrix(X)
if(nrow(X)==1) nc1=0 else nc1=ncol(X)
Z=as.matrix(Z)
if(nrow(Z)==1) nc2=0 else nc2=ncol(Z)
#nc1=ncol(X)  # should be same length as bvec (beta vector)
#nc2=ncol(Z)  # should be same length as gvec (gamma vector)
b0=param[1]
g0=param[(nc1+2)]
np=length(param)
rh1=param[np-1]; rh2=param[np]  # acf lag1 and lag2
cpar=c(rh1,rh2)
if(any(cpar<=-1) | any(cpar>=1)) return(1.e10)
dt=1+2*rh1*rh1*rh2-2*rh1*rh1-rh2*rh2
if(dt<=0) return(1.e10)
if(nc1>0)
{ bvec=param[2:(nc1+1)];
muvec=b0+X%*%bvec; muvec=exp(muvec)
}
else muvec=rep(exp(b0),n)
if(nc2>0)
{ gvec=param[(nc1+3):(nc1+nc2+2)];
omegavec=g0+Z%*%gvec; omegavec=exp(omegavec)/(1+exp(omegavec))
}
else omegavec=rep(exp(g0)/(1+exp(g0)),n)
#thv=(muvec, omegavec)
cdf1=pzipois(y,lambda=muvec,pstr0=omegavec)
pmf=dzipois(y,lambda=muvec,pstr0=omegavec)
cdf1[cdf1>=1]=1-1.e-9
cdf0=cdf1-pmf
cdf0[cdf0<=0]=1.e-9
z1=qnorm(cdf1); z0=qnorm(cdf0)  # convert to z values
# first probability
rmat12=matrix(c(1,rh1,rh1,1),2,2)
pmf12=pmvnorm(lower=z0[1:2],upper=z1[1:2],rep(0,2),rmat12)
nllk=-log(pmf12)
n=length(y)
rmat123=matrix(c(1,rh1,rh2,rh1,1,rh1,rh2,rh1,1),3,3)
for(i in 3:n)
{ ii=(i-2):i
tem=pmvnorm(lower=z0[ii], upper=z1[ii], rep(0,3),rmat123)
tem=tem
condpr=tem/pmf12
if(condpr<=0. | is.na(condpr)) condpr=1.e-15
nllk=nllk-log(condpr)
ii=(i-1):i
pmf12=pmvnorm(lower=z0[ii],upper=z1[ii],rep(0,2),rmat12)
}
nllk
}
```

REFERENCES

Adke, S., and S. Deshmukh, 1988. Limit distribution of a high order markov chain. *Journal of the Royal Statistical Society Series B (Methodological)* 105–108.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE transactions on automatic control* 19 (6):716–723.

Al-Osh, M., and A. A. Alzaid, 1987. First-order integer-valued autoregressive (inar (1)) process. *Journal of Time Series Analysis* 8 (3):261–275.

Al-Osh, M., and A. Alzaid, 1988. Integer-valued moving average (inma) process. *Statistical Papers* 29 (1):281–300.

Al-Osh, M., and A. Alzaid, 1991. Binomial autoregressive moving average models. *Stochastic Models* 7 (2):261–282.

Alqawba, M., N. Diawara, and J. M. Kim, 2019. Copula directional dependence of discrete time series marginals, to appear in Communications in Statistics - Simulation and Computation.

Alzaid, A., and M. Al-Osh, 1988. First-order integer-valued autoregressive (inar (1)) process: distributional and regression properties. *Statistica Neerlandica* 42 (1):53–61.

Alzaid, A., and M. Al-Osh, 1990. An integer-valued pth-order autoregressive structure (inar (p)) process. *Journal of Applied Probability* 27 (2):314–324.

Alzaid, A., and M. Al-Osh, 1993. Some autoregressive moving average processes with generalized poisson marginal distributions. *Annals of the Institute of Statistical Mathematics* 45 (2):223–232.

Andrews, D. W., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society* 817–858.

Billingsley, P., 1961. *Statistical inference for Markov processes*, volume 2. University of Chicago Press.

Börsch-Supan, A., and V. A. Hajivassiliou, 1993. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of econometrics* 58 (3):347–368.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel, 1994. *Time series analysis forecasting and control*. 3rd edition, Englewood Cliffs, N.J Prentice Hall.

Brockwell, P. J., and R. A. Davis, 2013. *Time series: theory and methods*. Springer Science & Business Media.

Buishand, T. A., 1978. The binary darma (1,1) process as a model for wet-dry sequences. Technical report, Dept Mathematics, Statistics Division, Agricultural University, Wageningen, Netherlands., 33, ,.

Buu, A., R. Li, X. Tan, and R. A. Zucker, 2012. Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Statistics in medicine* 31 (29):4074–4086.

Chang, T. J., J. Delleur, and M. Kavvas, 1987. Application of discrete autoregressive moving average models for estimation of daily runoff. *Journal of hydrology* 91 (1-2):119–135.

Chang, T. J., M. Kavvas, and J. Delleur, 1984a. Daily precipitation modeling by discrete autoregressive moving average processes. *Water Resources Research* 20 (5):565–580.

Chang, T. J., M. Kavvas, and J. Delleur, 1984b. Modeling of sequences of wet and dry days by binary discrete autoregressive moving average processes. *Journal of climate and applied meteorology* 23 (9):1367–1378.

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

Chen, C. W., M. K. So, J. C. Li, and S. Sriboonchitta, 2016. Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Statistical Methodology* 31:73–90.

Choo-Wosoba, H., and S. Datta, 2018. Analyzing clustered count data with a cluster-specific random effect zero-inflated conway–maxwell–poisson distribution. *Journal of Applied Statistics* 45 (5):799–814.

Conway, R. W., and W. L. Maxwell, 1962. A queuing model with state dependent service rates. *Journal of Industrial Engineering* 12 (2):132–136.

Cox, D., and H. Miller, 1965. *The theory of stochastic processes*. (Methuen, London, 1965).

Darsow, W. F., B. Nguyen, E. T. Olsen, et al., 1992. Copulas and markov processes. *Illinois journal of mathematics* 36 (4):600–642.

Davis, R. A., S. H. Holan, R. Lund, and N. Ravishanker, 2016. *Handbook of discrete-valued time series*. CRC Press.

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

Delleur, J., T. Chang, and M. Kavvas, 1989. Simulation models of sequences of dry and wet days. *Journal of irrigation and drainage engineering* 115 (3):344–357.

Dias, A., P. Embrechts, et al., 2004. Dynamic copula models for multivariate high-frequency data in finance. *Manuscript, ETH Zurich* 81.

Dunn, P. K., and G. K. Smyth, 1996. Randomized quantile residuals. *Journal of Computational and Graphical Statistics* 5 (3):236–244.

Fokianos, K., 2015. Statistical analysis of count time series models: A glm perspective. *Handbook of discrete-valued time series, Handbooks of Modern Statistical Methods* 3–28.

Frank, M. J., 1979. On the simultaneous associativity off (x, y) andx+y- f (x, y). *Aequationes mathematicae* 19 (1):194–226.

Genz, A., 1992. Numerical computation of multivariate normal probabilities. *Journal of computational and graphical statistics* 1 (2):141–149.

Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities.

Gonçalves, E., N. Mendes-Lopes, and F. Silva, 2016. Zero-inflated compound poisson distributions in integer-valued garch models. *Statistics* 50 (3):558–578.

Goodman, L. A., and W. H. Kruskal, 1954. Measures ofassociation for cross classifications. *Journal of the American Statistical Association* 49:732–64.

Goudie, A. S., and N. J. Middleton, 2006. *Desert dust in the global system*. Springer Science & Business Media.

Gourieroux, C., and A. Monfort, 1990. Simulation based inference in models with heterogeneity. *Annales d'Economie et de Statistique* 69–107.

Granger, C. W., 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* 424–438.

Gumbel, E. J., 1960. Distributions des valeurs extremes en plusiers dimensions. *Publ Inst Statist Univ Paris* 9:171–173.

Guolo, A., and C. Varin, 2014. Beta regression for time series analysis of bounded data, with application to canada google® flu trends. *The Annals of Applied Statistics* 8 (1):74–88.

Hajivassiliou, V., D. McFadden, and P. Ruud, 1996. Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of econometrics* 72 (1-2):85–134.

Hall, D. B., and Z. Zhang, 2004. Marginal models for zero inflated clustered data. *Statistical Modelling* 4 (3):161–180.

Hasan, M. T., G. Sneddon, and R. Ma, 2012. Regression analysis of zero-inflated time-series counts: application to air pollution related emergency room visit data. *Journal of Applied Statistics* 39 (3):467–476.

Ho, C.-H., and M. Bhaduri, 2015. On a novel approach to forecast sparse rare events: applications to parkfield earthquake prediction. *Natural Hazards* 78 (1):669–679.

Hothorn, T., F. Bretz, and A. Genz, 2001. On multivariate t and gauss probabilities in r. *sigma* 1000:3.

Ibragimov, R., 2009. Copula-based characterizations for higher order markov processes. *Econometric Theory* 25 (3):819–846.

Idso, S. B., 1976. Dust storms. *Scientific American* 235 (4):108–115.

Jacobs, P. A., and P. A. Lewis, 1978a. Discrete time series generated by mixtures. i: Correlational and runs properties. *Journal of the Royal Statistical Society: Series B (Methodological)* 40 (1):94–105.

Jacobs, P. A., and P. A. Lewis, 1978b. Discrete time series generated by mixtures ii: asymptotic properties. *Journal of the Royal Statistical Society: Series B (Methodological)* 40 (2):222–228.

Jacobs, P. A., and P. A. Lewis, 1978c. Discrete time series generated by mixtures. iii. autoregressive processes (dar (p)). Technical report, NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF.

Jia, Y., S. Kechagias, J. Livsey, R. Lund, and V. Pipiras, 2018. Latent gaussian count time series modeling. *arXiv preprint arXiv:181100203* .

Joe, H., 1995. Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association* 90 (431):957–964.

Joe, H., 1996. Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability* 33 (3):664–677.

Joe, H., 1997. *Multivariate models and multivariate dependence concepts*. Chapman and Hall/CRC.

Joe, H., 2014. *Dependence modeling with copulas*. Chapman and Hall/CRC.

Joe, H., 2016. Markov models for count time series. In *Handbook of Discrete-Valued Time Series*, 49–70, Chapman and Hall/CRC.

Joe, H., and T. Hu, 1996. Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of multivariate analysis* 57 (2):240–265.

Jung, R. C., and A. Tremayne, 2011. Convolution-closed models for count time series with applications. *Journal of Time Series Analysis* 32 (3):268–280.

Keane, M. P., 1994. A computationally practical simulation estimator for panel data. *Econometrica: Journal of the Econometric Society* 95–116.

Kedem, B., and K. Fokianos, 2005. *Regression models for time series analysis*, volume 488. John Wiley & Sons.

Kim, J. M., and S. Hwang, 2017. Directional dependence via gaussian copula beta regression model with asymmetric garch marginals. *Communications in Statistics-Simulation and Computation* 46 (10):7639–7653.

Lahiri, S. N., 2013. *Resampling methods for dependent data*. Springer Science & Business Media.

Lambert, D., 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics* 34 (1):1–14.

Lee, L.-f., 1999. Statistical inference with simulated likelihood functions. *Econometric Theory* 15 (3):337–360.

Lennon, H., 2016. Gaussian copula modelling for integer-valued time series. Ph.D. thesis, University of Manchester.

Lennon, H., and J. Yuan, 2019. Estimation of a digitised gaussian arma model by monte carlo expectation maximisation. *Computational Statistics & Data Analysis* 133:277–284.

MacDonald, I. L., and W. Zucchini, 1997. *Hidden Markov and other models for discrete-valued time series*, volume 110. CRC Press.

Mardia, K. V., 1967. Some contributions to contingency-type bivariate distributions. *Biometrika* 54 (1-2):235–249.

Masarotto, G., and C. Varin, 2012. Gaussian copula marginal regression. *Electronic Journal of Statistics* 6:1517–1549.

Masarotto, G., and C. Varin, 2017. Gaussian copula regression in r. *Journal of Statistical Software* 77 (8):1–26.

McKenzie, E., 1985. Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association* 21 (4):645–650.

McKenzie, E., 1986. Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability* 18 (3):679–705.

McKenzie, E., 1987. Innovation distributions for gamma and negative binomial autoregressions. *Scandinavian journal of statistics* 79–85.

McKenzie, E., 1988. Some arma models for dependent sequences of poisson counts. *Advances in Applied Probability* 20 (4):822–835.

McKenzie, E., 2003. Ch. 16. discrete variate time series. *Handbook of statistics* 21:573–606.

Nelder, J. A., and R. W. M. Wedderburn, 1972. Generalized linear models. *Journal of the Royal Statistical Society Series A (General)* 135 (3):370–384.

Nelsen, R. B., 2007. *An introduction to copulas*. Springer.

Nešlehová, J., 2007. On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis* 98 (3):544–567.

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.

Nikoloulopoulos, A. K., and D. Karlis, 2010. Regression in a copula model for bivariate count data. *Journal of Applied Statistics* 37 (9):1555–1568.

Nikoloulopoulos, A. K., and P. G. Moffatt, 2019. Coupling couples with copulas: analysis of assortative matching on risk attitude. *Economic Inquiry* 57 (1):654–666.

Palaro, H. P., and L. K. Hotta, 2006. Using conditional copula to estimate value at risk. *Journal of Data Science* 4:93–115.

Panagiotelis, A., C. Czado, and H. Joe, 2012. Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association* 107 (499):1063–1072.

Patton, A. J., 2009. Copula-based models for financial time series. In *Handbook of financial time series*, 767–785, Springer.

Pipiras, V., and M. S. Taqqu, 2017. *Long-range dependence and self-similarity*, volume 45. Cambridge university press.

Pitt, M., D. Chan, and R. Kohn, 2006. Efficient bayesian inference for gaussian copula regression models. *Biometrika* 93 (3):537–554.

Plackett, R. L., 1965. A class of bivariate distributions. *Journal of the American Statistical Association* 60 (310):516–522.

Politis, D. N., and J. P. Romano, 1994. The stationary bootstrap. *Journal of the American Statistical association* 89 (428):1303–1313.

R Core Team, 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A. E., 1985. A model for high-order markov chains. *Journal of the Royal Statistical Society Series B (Methodological)* 528–539.

Ridout, M., J. Hinde, and C. G. DeméAtrio, 2001. A score test for testing a zero-inflated poisson regression model against zero-inflated negative binomial alternatives. *Biometrics* 57 (1):219–223.

Sellers, K. F., and A. Raim, 2016. A flexible zero-inflated model to address data dispersion. *Computational Statistics & Data Analysis* 99:68–80.

Serfozo, R., 2009. *Basics of applied stochastic processes*. Springer Science & Business Media.

Shmueli, G., T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright, 2005. A useful distribution for fitting discrete data: revival of the conway–maxwell–poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54 (1):127–142.

Shumway, R. H., and D. S. Stoffer, 2011. Time series regression and exploratory data analysis. In *Time series analysis and its applications*, 47–82, Springer.

Steutel, F., and K. Van Harn, 1979. Discrete analogues of self-decomposability and stability. *The Annals of Probability* 893–899.

Sungur, E. A., 2005. A note on directional dependence in regression setting. *Communications in Statistics—Theory and Methods* 34 (9-10):1957–1965.

Tan, S., M. Bhaduri, and C.-H. Ho, 2014. A statistical model for long-term forecasts of strong sand dust storms. *Journal of Geoscience and Environment Protection* 2 (03):16.

Tao, G., L. Jingtao, Y. Xiao, K. Ling, F. Yida, and H. Yinghua, 2002. Objective pattern discrimination model for dust storm forecasting. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling* 9 (1):55–62.

Weiß, C. H., 2008. Thinning operations for modeling time series of counts—a survey. *AStA Advances in Statistical Analysis* 92 (3):319.

Yang, M., J. E. Cavanaugh, and G. K. Zamba, 2015. State-space models for count time series with excess zeros. *Statistical Modelling* 15 (1):70–90.

Yang, M., G. K. Zamba, and J. E. Cavanaugh, 2013. Markov regression models for count time series with excess zeros: A partial likelihood approach. *Statistical Methodology* 14:26–38.

Yau, K. K., A. H. Lee, and P. J. Carrivick, 2004. Modeling zero-inflated count series with application to occupational health. *Computer methods and programs in biomedicine* 74 (1):47–52.

Yau, K. K., K. Wang, and A. H. Lee, 2003. Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 45 (4):437–452.

Zhu, F., 2011. A negative binomial integer-valued garch model. *Journal of Time Series Analysis* 32 (1):54–67.

Zhu, F., 2012. Zero-inflated poisson and negative binomial integer-valued garch models. *Journal of Statistical Planning and Inference* 142 (4):826–839.

Zhu, R., and H. Joe, 2006. Modelling count data time series with markov processes based on binomial thinning. *Journal of Time Series Analysis* 27 (5):725–738.

Zhu, R., and H. Joe, 2010a. Count data time series models based on expectation thinning. *Stochastic Models* 26 (3):431–462.

Zhu, R., and H. Joe, 2010b. Negative binomial time series models based on expectation thinning operators. *Journal of Statistical Planning and Inference* 140 (7):1874–1888.

VITA

Mohammed Sulaiman Alqawba
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

Education

Ph.D.	Old Dominion University, Norfolk, VA, USA	August 2019
	Major: Computational and Applied Mathematics (Statistics).	
M.S.	Old Dominion University, Norfolk, VA, USA	May 2016
	Major: Computational and Applied Mathematics (Statistics).	
B.S.	Qassim University, Alrass, Saudi Arabia	July 2011
	Major: Mathematics.	

Experience

Graduate Research Assistant, Old Dominion University	May 2019 - Aug 2019
Virginia Modeling, Analysis & Simulation Center, Suffolk, VA, USA	
Graduate Research Assistant, Old Dominion University	May 2018 - Aug 2018
Department of Mathematics and Statistics, Norfolk, VA, USA	
Teaching Assistant, Qassim University	Aug 2013 - October 2013
Department of Mathematics, Alrass, Saudi Arabia	
Mathematics Teacher, Ministry of Education	Aug 2011 - May 2013
Education Administration, Alrass, Saudi Arabia	

Typeset using LaTeX.