VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Magistro darbas

JUNGČIŲ TAIKYMAS PASKOLŲ MODELIAVIME

APPLICATIONS OF COPULAS IN LOAN MODELLING

Andrius Buteikis

VILNIUS 2016

MATEMATIKOS IR INFORMATIKOS FAKULTETAS
EKONOMETRINĖS ANALIZĖS KATEDRA

Darbo vadovas Prof. Habil. Dr. Remigijus Leipus

Darbo recenzentas Dr. Jurgita Markevičiūtė

Darbas apgintas 2016 m. sausio 14 d.
Darbas įvertintas ____________________

Registravimo Nr. 111000-9.1-5/_______
2016-01-04 _________________

Contents

Jungčių taikymas paskolų modeliavime

Santrauka

Jungčių taikymas diskretiems duomenims, pasižymintiems priklausomumu, nėra plačiai išnagrinėtas literatūroje. Šiame darbe nagrinėjamas pirmos eilės dvimatis sveikareikšmių dydžių autoregresinis procesas (BINAR(1)), kurio paklaidos yra aprašomos jungties funkcija. Pateikiamos modelio savybės su įrodymais. Nagrinėjami skirtingi vertinimo metodai, o jų palyginimai atliekami Monte Karlo simuliacijomis, pabrėžiant jungties priklausomybės parametro vertinimo galimybes. Empiriniams mokių ir nemokių paskolų dieniniams duomenims sudaromi BINAR(1) modeliai. Naudojamos skirtingos jungčių funkcijų ir marginaliųjų pasiskirstymo funkcijų kombinacijos. Vertinami modeliai su liekanų marginaliosiomis funkcijomis, kurios gali būti iš tos pačios arba iš skirtingų skirstinių šeimų.

Raktiniai žodžiai : Skaičiuojantieji duomenys, BINAR, Puasono, Neigiamas binominis skirstinys, jungtis, FGM jungtis, Frank jungtis, Clayton jungtis, Gumbel jungtis

Applications of copulas in loan modelling

Abstract

Copula applications for discrete data with autocorrelation are not widely studied. In this thesis, a bivariate integer-valued autoregressive process of order 1 (BINAR(1)) with copula-joint innovations is analysed. Model properties and their proofs are provided. Different estimation methods are analysed and comparisons are carried out via Monte Carlo simulations with emphasis on estimation of the copula dependence parameter. An empirical application on defaulted and non-defaulted loan daily data is carried out using different combinations of copula functions and marginal distribution functions covering the cases when both marginal distributions are from the same family and when they are from different distribution families.

Key words : Count data, BINAR, Poisson, Negative binomial distribution, Copula, FGM copula, Frank copula, Clayton copula, Gumbel copula

1 Introduction

Copulas are functions which link marginal distributions of a random vector to form a joint distribution. The advantage is that copulas allow to model the marginal distributions (which can be from different distribution families) and their dependence structure (which is described via a copula) separately. Because of this feature copulas were applied to many different fields, including finance[1] (for examples of copula applications see Brigo et al. (2010) or Cherubini et al. (2011)), which also included the analysis of loans and their default rates. Crook and Moreira (2011) analysed the dependence between the default rate of loans between different credit risk categories. In order to model the dependence, copulas from 10 different families were applied and three model selection tests were carried out. Because of the small sample size (24 observations per risk category) most of the copula families were not rejected and a single best copula model was not selected. To analyse whether dependence is affected by time Fenech et al. (2015) estimated the dependence between four different loan default indexes before the Great Financial Crisis and after. They have found that the dependence was different in these periods. Four copula families were used to estimate the dependence between the default index pairs.

While these studies were carried out for continuous data, there is less developed literature on discrete models created with copulas: Genest and Nešlehová (2007) discusses the differences and challenges of using copulas for discrete data compared to continuous data. Furthermore, the previously mentioned studies assumed that the data does not depend on its own previous values. By using bivariate integer-valued autoregressive models (BINAR) it is possible to account for both the discreteness and autocorrelation of the data. Furthermore, copulas can be used to model the dependence of innovations in the BINAR(1) models: Karlis and Pedeli (2013) used the Frank copula and normal copula to model the dependence of the innovations of the BINAR(1) model.

In this thesis we expand on using copulas in BINAR models by analysing additional copula families for the innovations of the BINAR(1) model. Secondly, we analysed different estimation methods for BINAR(1) models. We also present a two-stage estimation method for the parameters of the BINAR(1) model where we estimate the model parameters separately from the dependence parameter of the copula. These estimation methods are then compared to the estimation method used in Karlis and Pedeli (2013) via Monte Carlo simulations. Additionally, an estimation method for the dependence parameter of a copula with discrete marginal distributions when the data does not exhibit autoregressive properties is examined in this thesis. This method is based on minimizing the distance between the theoretical copula and the empirical joint distribution function. It is then compared to one of the more commonly used estimation methods for copula dependence parameter estimation via Monte Carlo simulations. Finally, in order to analyse the presence of autocorrelation and copula dependence in loan data, an empirical application is carried out.

The thesis is structured as follows: Section 2 presents the BINAR(1) process and its main properties, Section 3 presents the main properties of copulas as well as some copula functions with graphical examples. Section 4 compares different estimation methods for the BINAR(1) model and the dependence parameter of copulas via Monte Carlo simulations. In Section 5 an empirical application is carried out using different combinations of copula functions and marginal distribution functions. Conclusions are presented in Section 6.

[1] other notable fields include survival analysis, hydrology and insurance risk analysis

2 The bivariate INAR(1) process

The BINAR(1) process was introduced in Pedeli and Karlis (2011). We will begin this Section by providing the definition of the BINAR(1) model as well as its properties.

2.1 The BINAR(1) model

Definition 2.1. Let $\mathbf{X_t}$, $t \in \mathbb{Z}$ be a non-negative integer-valued bivariate time series and let $\mathbf{R_t}$, $t \in \mathbb{Z}$ be a non-negative integer-valued bivariate random sequence. Then a bivariate integer-valued autoregressive process of order 1 (BINAR(1)) is defined as:

$$\mathbf{X_t} = \mathbf{A} \circ \mathbf{X_{t-1}} + \mathbf{R_t} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} R_{1,t} \\ R_{2,t} \end{bmatrix}, \quad t \in \mathbb{Z}. \tag{2.1}$$

The symbol '$\circ$' is the thinning operator which also acts as the matrix multiplication. So the $j = 1, 2$ element is defined as an INAR process of order 1 (INAR(1)):

$$X_{j,t} = \alpha_j \circ X_{j,t-1} + R_{j,t}, \tag{2.2}$$

where $\alpha \circ X := \sum_{i=1}^{X} Y_i$ and $Y_1, Y_2, ...$ is a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(Y_i = 1) = \alpha = 1 - \mathbb{P}(Y_i = 0)$, $\alpha \in [0, 1)$. $X_{j,t}$ has two random components: the survivors of the elements of the process at time $t-1$, each with the probability of survival α_j, which is denoted by $\alpha_j \circ X_{j,t-1}$, and the elements which enter in the system in the interval $[t-1, t]$ which are called arrival elements $R_{j,t}$. The distribution properties of the BINAR(1) process can be studied in terms of $\mathbf{R}_t$ values. We can obtain a moving average representation by substitutions and the properties of the thinning operator as in Kedem and Fokianos (2002):

$$\begin{aligned} X_{j,t} &= \alpha_j \circ X_{j,t-1} + R_{j,t} \\ &= \alpha_j \circ (\alpha_j \circ X_{j,t-2} + R_{j,t-1}) + R_{j,t} \\ &\stackrel{d}{=} \alpha_j^2 \circ X_{j,t-2} + \alpha_j \circ R_{j,t-1} + R_{j,t} \\ &\stackrel{d}{=} ... \\ &\stackrel{d}{=} \sum_{k=0}^{\infty} \alpha_j^k \circ R_{j,t-k}, \quad j = 1, 2, \quad t \in \mathbb{Z}. \end{aligned} \tag{2.3}$$

Pedeli (2011) states that $\mathbf{X_t}$, $t \in \mathbb{Z}$ is strictly stationary, if the largest eigenvalue of the matrix $\mathbf{A}$ is less than 1, i.e. if $\max(\alpha_1, \alpha_2) < 1$ and $R_{j,t}$ - an i.i.d. sequence with $\mathbb{E}R_{j,t} < \infty$, $j = 1, 2$. A number of thinning operator properties are provided in Pedeli (2011) and Silva (2005) with proofs for selected few. We will present the main properties of the thinning operator along with their proofs which will be used later on for the BINAR(1) model property validity:

Theorem 2.1. ***Thinning operator properties***. *Let X, X_1, X_2 be non-negative integer-valued random variables, $\alpha, \alpha_1, \alpha_2 \in [0, 1)$ and let '$\circ$' be the thinning operator. Then the following properties hold:*

(a) $0 \circ X = 0$, $1 \circ X = X$;

(b) $\alpha_1 \circ (\alpha_2 \circ X) \stackrel{d}{=} (\alpha_1\alpha_2) \circ X$;

(c) $\alpha \circ (X_1 + X_2) \stackrel{d}{=} \alpha \circ X_1 + \alpha \circ X_2$;

(d) $\mathbb{E}(\alpha \circ X) = \alpha\mathbb{E}(X)$;

(e) $\mathbb{V}\text{ar}(\alpha \circ X) = \alpha^2 \mathbb{V}\text{ar}(X) + \alpha(1-\alpha)\mathbb{E}(X)$;

(f) $\mathbb{E}((\alpha \circ X_1)X_2) = \alpha \mathbb{E}(X_1 X_2)$;

(g) $\mathbb{C}\text{ov}(\alpha \circ X_1, X_2) = \alpha \mathbb{C}\text{ov}(X_1, X_2)$;

(h) $\mathbb{E}((\alpha_1 \circ X_1)(\alpha_2 \circ X_2)) = \alpha_1 \alpha_2 \mathbb{E}(X_1 X_2)$.

Proof:

(a) $0 \circ X = \sum_{i=1}^{X} Y_i \overset{a.s.}{=} \sum_{i=1}^{X} 0 = 0$, since $\mathbb{P}(Y_i = 1) = 0$ and $\mathbb{P}(Y_i = 0) = 1$, $i = 1, 2, ...$; $1 \circ X = \sum_{i=1}^{X} Y_i \overset{a.s.}{=} \sum_{i=1}^{X} 1 = X$, since $\mathbb{P}(Y_i = 1) = 1$ and $\mathbb{P}(Y_i = 0) = 0$, $i = 1, 2, ...$;

(b) Let $\widetilde{Y}_i \sim Bern(\alpha_1)$, $Y_i \sim Bern(\alpha_2)$ and $\overline{Y}_i \sim Bern(\alpha_1\alpha_2)$, $i = 1, 2, ...$. We will show that $\alpha_1 \circ (\alpha_2 \circ X)$ and $(\alpha_1\alpha_2) \circ X$ are equal in distribution by showing that their characteristic functions are equal. From the definition of the thinning operator we have that

$$\alpha_1 \circ (\alpha_2 \circ X) = \alpha_1 \circ \sum_{i=1}^{X} Y_i = \sum_{j=1}^{\sum_{i=1}^{X} Y_i} \widetilde{Y}_j, \tag{2.4}$$

$$(\alpha_1\alpha_2) \circ X = \sum_{i=1}^{X} \overline{Y}_i. \tag{2.5}$$

The characteristic function of equation (2.4) is:

$$\begin{aligned}\overline{\varphi}(u) &= \mathbb{E}e^{iu\sum_{i=1}^{X} \overline{Y}_i} = \sum_{k=0}^{\infty} \mathbb{E}e^{iu\sum_{i=1}^{k} \overline{Y}_i}\mathbb{P}(X = k) = \sum_{k=0}^{\infty} \overline{\varphi}^k_{\alpha_1,\alpha_2}(u)\mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty} (e^{iu}\alpha_1\alpha_2 + (1 - \alpha_1\alpha_2))^k \mathbb{P}(X = k),\end{aligned} \tag{2.6}$$

because

$$\begin{aligned}\overline{\varphi}_{\alpha_1,\alpha_2}(u) &= \mathbb{E}e^{iu\overline{Y}} = \mathbb{E}(e^{iu\overline{Y}}|\overline{Y} = 1)\mathbb{P}(\overline{Y} = 1) + \mathbb{E}(e^{iu\overline{Y}}|\overline{Y} = 0)\mathbb{P}(\overline{Y} = 0) \\ &= e^{iu}\alpha_1\alpha_2 + (1 - \alpha_1\alpha_2).\end{aligned}$$

The characteristic function of equation (2.5) is:

$$\widetilde{\varphi}(u) = \mathbb{E}e^{iu\sum_{j=1}^{\sum_{i=1}^{X} Y_i} \widetilde{Y}_j} = \sum_{k=0}^{\infty} \mathbb{E}e^{iu\sum_{j=1}^{\sum_{i=1}^{k} Y_i} \widetilde{Y}_j}\mathbb{P}(X = k). \tag{2.7}$$

We see that

$$\mathbb{E}e^{iu\sum_{j=1}^{\sum_{i=1}^{k} Y_i} \widetilde{Y}_j} = \sum_{l=0}^{k} \mathbb{E}e^{iu\sum_{j=1}^{l} \widetilde{Y}_j}\mathbb{P}\left(\sum_{i=1}^{k} Y_i = l\right) = \sum_{l=0}^{k} \varphi^l_{\alpha_1}(u)\mathbb{P}\left(\sum_{i=1}^{k} Y_i = l\right), \tag{2.8}$$

where

$$\begin{aligned}\varphi_{\alpha_1}(u) &= \mathbb{E}e^{iu\widetilde{Y}} = \mathbb{E}(e^{iu\widetilde{Y}}|\widetilde{Y} = 1)\mathbb{P}(\widetilde{Y} = 1) + \mathbb{E}(e^{iu\widetilde{Y}}|\widetilde{Y} = 0)\mathbb{P}(\widetilde{Y} = 0) \\ &= e^{iu}\alpha_1 + (1 - \alpha_1),\end{aligned}$$

and

$$\mathbb{P}\left(\sum_{i=1}^{k} Y_i = l\right) = \binom{k}{l}\alpha_2^l(1-\alpha_2)^{k-l}.$$

From (2.7) and (2.8) we have that:

$$\widetilde{\varphi}(u) = \mathbb{E}e^{iu\sum_{j=1}^{\sum_{i=1}^{X} Y_i}\widetilde{Y}_j} = \sum_{k=0}^{\infty}\sum_{l=0}^{k}(e^{iu}\alpha_1 + (1-\alpha_1))^l\binom{k}{l}\alpha_2^l(1-\alpha_2)^{k-l}\mathbb{P}(X=k) \quad (2.9)$$

Because $(a+b)^k = \sum_{l=0}^{k}\binom{k}{l}a^l b^{k-l}$ where $a, b \in \mathbb{R}$, we see that $\sum_{l=0}^{k}(e^{iu}\alpha_1 + (1-\alpha_1))^l\binom{k}{l}\alpha_2^l(1-\alpha_2)^{k-l} = (e^{iu}\alpha_1\alpha_2 + (1-\alpha_1)\alpha_2 + (1-\alpha_2))^k$ and equation (2.9) becomes:

$$\begin{aligned}\widetilde{\varphi}(u) &= \sum_{k=0}^{\infty}(e^{iu}\alpha_1\alpha_2 + (1-\alpha_1)\alpha_2 + (1-\alpha_2))^k\mathbb{P}(X=k)\\ &= \sum_{k=0}^{\infty}(e^{iu}\alpha_1\alpha_2 + (1-\alpha_1\alpha_2))^k\mathbb{P}(X=k). \quad (2.10)\end{aligned}$$

Comparing equations (2.6) and (2.10) we see that they are equal, so we have that $\alpha_1 \circ (\alpha_2 \circ X) \stackrel{d}{=} (\alpha_1\alpha_2) \circ X$.

(c) We have that $\alpha \circ (X_1 + X_2) = \sum_{i=1}^{X_1+X_2} Y_i$, which has the characteristic function:

$$\begin{aligned}\varphi(u) = \mathbb{E}e^{iu\sum_{i=1}^{X_1+X_2} Y_i} &= \sum_{k=0}^{\infty}\mathbb{E}e^{iu\sum_{i=1}^{k} Y_i}\mathbb{P}(X_1+X_2=k) \quad (2.11)\\ &= \sum_{k=0}^{\infty}(e^{iu}\alpha + (1-\alpha))^k\mathbb{P}(X_1+X_2=k), \quad (2.12)\end{aligned}$$

where we used the property that Y_i are i.i.d. random variables:

$$\mathbb{E}e^{iu\sum_{i=1}^{k} Y_i} = (\mathbb{E}e^{iuY_i})^k = (e^{iu}\alpha + (1-\alpha))^k.$$

Applying the same propertis to the right side of the equality, we have that:

$$\alpha \circ X_1 + \alpha \circ X_2 = \sum_{i=1}^{X_1}\overline{Y}_i + \sum_{i=1}^{X_2}\widetilde{Y}_i = \sum_{i=1}^{X_1+X_2} Y_i,$$

where $\overline{Y}_i \sim Bern(\alpha)$ and $\widetilde{Y}_i \sim Bern(\alpha)$ and

$$Y_i = \begin{cases}\overline{Y}_i, & \text{if } i = 1, ..., X_1,\\ \widetilde{Y}_i, & \text{if } i = X_1+1, ..., X_2.\end{cases}$$

Since Y_i are i.i.d. random variables conditionally with respect to X_1 and X_2, we have that the characteristic function is the same as the left side of the equality:

$$\widetilde{\varphi}(u) = \mathbb{E}e^{iu\sum_{i=1}^{X_1+X_2} Y_i} = \sum_{k=0}^{\infty}(e^{iu}\alpha + (1-\alpha))^k\mathbb{P}(X_1+X_2=k) = \varphi(u). \quad (2.13)$$

Thus, the equality in part (c) holds.

(d) Using the definition of $\alpha \circ X$ we have that:

$$\begin{aligned}\mathbb{E}\left(\alpha \circ X\right) &= \mathbb{E}\left(\sum_{i=1}^{X} Y_i\right) = \sum_{k=0}^{\infty} \mathbb{E}\left(\sum_{i=1}^{k} Y_i\right) \mathbb{P}(X=k) \\ &= \sum_{k=0}^{\infty}\sum_{i=1}^{k} \alpha \mathbb{P}(X=k) = \alpha \sum_{k=0}^{\infty} k\mathbb{P}(X=k) = \alpha\mathbb{E}(X).\end{aligned}$$

(e) We know that $\mathbb{V}\text{ar}(\alpha \circ X) = \mathbb{E}(\alpha \circ X)^2 - (\mathbb{E}(\alpha \circ X))^2$. Using part (d), we have that the first term can be expressed as

$$(\mathbb{E}(\alpha \circ X))^2 = (\alpha\mathbb{E}(X))^2 = \alpha^2(\mathbb{E}(X))^2, \tag{2.14}$$

and the second term as

$$\begin{aligned}\mathbb{E}(\alpha \circ X)^2 &= \mathbb{E}\left(\sum_{i=1}^{X} Y_i\right)^2 = \mathbb{E}\left(\sum_{i,j=1}^{X} Y_iY_j\right) = \mathbb{E}\left(\sum_{i\neq j} Y_iY_j + \sum_{i=1}^{X} Y_i^2\right) \\ &= \sum_{k=0}^{\infty} \mathbb{E}\left(\sum_{i\neq j} Y_iY_j + \sum_{i=1}^{k} Y_i^2\right) \mathbb{P}(X=k) = \sum_{k=0}^{\infty}\left(\sum_{i\neq j} \alpha^2 + \sum_{i=1}^{k} \alpha\right)\mathbb{P}(X=k) \\ &= \sum_{k=0}^{\infty}((k^2-k)\alpha^2 + k\alpha)\mathbb{P}(X=k) \\ &= \alpha^2\sum_{k=0}^{\infty} k^2\mathbb{P}(X=k) + \alpha(1-\alpha)\sum_{k=0}^{\infty} k\mathbb{P}(X=k) \\ &= \alpha^2\mathbb{E}(X^2) + \alpha(1-\alpha)\mathbb{E}(X).\end{aligned} \tag{2.15}$$

From equations (2.14) and (2.15) we get:

$$\begin{aligned}\mathbb{V}\text{ar}(\alpha \circ X) &= \alpha^2\mathbb{E}(X^2) + \alpha(1-\alpha)\mathbb{E}(X) - \alpha^2(\mathbb{E}(X))^2 \\ &= \alpha^2\mathbb{V}\text{ar}(X) + \alpha(1-\alpha)\mathbb{E}(X).\end{aligned}$$

(f) We have that

$$\begin{aligned}\mathbb{E}((\alpha \circ X_1)X_2) &= \mathbb{E}\left(\mathbb{E}((\alpha \circ X_1)X_2|X_1,X_2)\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{X_1} Y_iX_2|X_1,X_2\right)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{X_1}\mathbb{E}\left(Y_iX_2|X_1,X_2\right)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{X_1}\mathbb{E}\left(Y_i|X_1,X_2\right)X_2\right) \\ &= \mathbb{E}\left(X_2\sum_{i=1}^{X_1}\alpha\right) \\ &= \alpha\mathbb{E}(X_1X_2).\end{aligned} \tag{2.16}$$

(g) Using parts (d) and (f), we get that

$$\begin{aligned}\mathbb{C}\text{ov}(\alpha \circ X_1, X_2) &= \mathbb{E}((\alpha \circ X_1)X_2) - \mathbb{E}(\alpha \circ X_1)\mathbb{E}(X_2) \\ &= \alpha\mathbb{E}(X_1X_2) - \alpha\mathbb{E}(X_1)\mathbb{E}(X_2) \\ &= \alpha\left(\mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)\right) \\ &= \alpha\mathbb{C}\text{ov}(X_1, X_2).\end{aligned} \tag{2.17}$$

(h) Similarly to part (f), we have that

$$\begin{aligned}\mathbb{E}((\alpha_1 \circ X_1)(\alpha_2 \circ X_2)) &= \mathbb{E}\left(\mathbb{E}((\alpha \circ X_1)(\alpha_2 \circ X_2)|X_1, X_2)\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{X_1}\sum_{j=1}^{X_2} Y_iY_j|X_1, X_2\right)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{X_1}\sum_{j=1}^{X_2} \mathbb{E}\left(Y_iY_j|X_1, X_2\right)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{X_1}\sum_{j=1}^{X_2} \mathbb{E}\left(Y_i|X_1, X_2\right)\mathbb{E}\left(Y_j|X_1, X_2\right)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{X_1}\sum_{j=1}^{X_2} \alpha_1\alpha_2\right) \\ &= \alpha_1\alpha_2\mathbb{E}(X_1X_2).\end{aligned} \tag{2.18}$$

2.2 The properties of the BINAR(1) model

A number of the properties of the BINAR(1) model are presented in Pedeli (2011) with proofs for a selected few. For the convenience of the reader in this section we will present the properties of the BINAR(1) process (2.1) along with their complete proofs. These properties will later be used when analysing some of the parameter estimation methods.

Theorem 2.2. ***Properties of the BINAR(1) process.*** *Let $\mathbf{X}_t = (X_{1,t}, X_{2,t})'$ be a non-negative integer-valued time series given in Def. 2.1 and $\alpha_j \in [0, 1)$, $j = 1, 2$. Let $\mathbf{R}_t = (R_{1,t}, R_{2,t})'$ be a non negative integer-valued random variables with a finite mean λ_j and variance σ_j^2, $j = 1, 2$. Then the following properties hold:*

(a) $\mathbb{E}X_{j,t} = \mu_{X_j} = \frac{\lambda_j}{1-\alpha_j}$;

(b) $\mathbb{E}(X_{j,t}|X_{j,t-1}) = \alpha_j X_{j,t-1} + \lambda_j$;

(c) $\mathbb{V}\text{ar}(X_{j,t}) = \sigma^2_{X_j} = \dfrac{\sigma_j^2 + \alpha_j\lambda_j}{1-\alpha_j}$;

(d) $\mathbb{C}\text{ov}(X_{i,t}, R_{j,t}) = \mathbb{C}\text{ov}(R_{i,t}, R_{j,t})$, $i \neq j$;

(e) $\mathbb{C}\text{ov}(X_{j,t}, X_{j,t+h}) = \alpha_j^h \sigma^2_{X_j}$;

(f) $\mathbb{C}\text{orr}(X_{j,t}, X_{j,t+h}) = \alpha_j^h$, $h \geq 0$;

(g) $\mathbb{C}\text{ov}(X_{i,t}, X_{j,t+h}) = \dfrac{\alpha_j^h}{1-\alpha_i\alpha_j}\mathbb{C}\text{ov}(R_{i,t}, R_{j,t})$, $i \neq j$, $h \geq 0$;

(h) $\mathbb{C}\text{orr}(X_{i,t+h}, X_{j,t}) = \dfrac{\alpha_i^h\sqrt{(1-\alpha_i^2)(1-\alpha_j^2)}}{(1-\alpha_1\alpha_2)\sqrt{(\sigma_i^2+\alpha_i\lambda_i)(\sigma_j^2+\alpha_j\lambda_j)}}\mathbb{C}\text{ov}(R_{i,t}, R_{j,t}),\ i \neq j,\ h \geq 0;$

Proof:

(a) We have

$$\begin{aligned}
\mathbb{E}X_{j,t} &= \mathbb{E}(\alpha_j \circ X_{j,t-1} + R_{j,t}) = \mathbb{E}(\alpha_j^2 \circ X_{j,t-2} + \alpha_j \circ R_{j,t-1} + R_{j,t}) \\
&= \ldots = \mathbb{E}(\sum_{k=0}^{\infty} \alpha_j^k \circ R_{j,t-k}) = \sum_{k=0}^{\infty} \alpha_j^k \mathbb{E}(R_{j,t-k}) = \sum_{k=0}^{\infty} \alpha_j^k \lambda_j \\
&= \frac{\lambda_j}{1-\alpha_j}.
\end{aligned}$$

Here, first equality is from the definition of BINAR(1) model from equation (2.1). We get the second, third and fourth equalities by using the definition of BINAR(1) model expressed in terms of arrival processes (2.3) and the second and third properties from Theorem 2.1. The fifth equality is from the fourth property from Theorem 2.1. The last equality is from the definition of an infinite geometric series.

(b) We have

$$\begin{aligned}
\mathbb{E}(X_{j,t}|X_{j,t-1}) &= \mathbb{E}(\alpha_j \circ X_{j,t-1} + R_{j,t}|X_{j,t-1}) = \mathbb{E}(\alpha_j \circ X_{j,t-1}|X_{j,t-1}) + \mathbb{E}(R_{j,t}|X_{j,t-1}) \\
&= \alpha_j \mathbb{E}(X_{j,t-1}|X_{j,t-1}) + \mathbb{E}(R_{j,t}) = \alpha_j X_{j,t-1} + \lambda_j.
\end{aligned}$$

(c) We have

$$\begin{aligned}
\mathbb{V}\text{ar}(X_{j,t}) &= \mathbb{V}\text{ar}\left(\sum_{k=0}^{\infty} \alpha_j^k \circ R_{j,t-k}\right) = \sum_{k=0}^{\infty} \mathbb{V}\text{ar}(\alpha_j^k \circ R_{j,t-k}) \\
&= \sum_{k=0}^{\infty} \left(\alpha_j^{2k}\mathbb{V}\text{ar}(R_{j,t-k}) + \alpha_j^k(1-\alpha_j^k)\mathbb{E}(R_{j,t-k})\right) \\
&= \sum_{k=0}^{\infty}(\alpha_j^{2k}\sigma_j^2 + \alpha_j^k(1-\alpha_j^k)\lambda_j) = \frac{1}{1-\alpha_j^2}\sigma_j^2 + \frac{1}{1-\alpha_j}\lambda_j - \frac{1}{1-\alpha_j^2}\lambda_j \\
&= \frac{\sigma_j^2 + \lambda_j + \alpha_j\lambda_j - \lambda_j}{1-\alpha_j^2} = \frac{\sigma_j^2 + \alpha_j\lambda_j}{1-\alpha_j^2}.
\end{aligned}$$

Here, the first equality is from equation (2.3). The second equality is from the fact that $R_{j,t-k}$ are i.i.d. The third equality is from the fifth property of Theorem 2.1.

(d) We have that

$$\begin{aligned}
\mathbb{C}\text{ov}(X_{i,t}, R_{j,t}) &= \mathbb{E}(X_{i,t}R_{j,t}) - \mathbb{E}(X_{i,t})\mathbb{E}(R_{j,t}) = \mathbb{E}(X_{i,t}R_{j,t}) - \mu_{X_i}\lambda_j \\
&= \mathbb{E}\left(\left(\sum_{k=0}^{\infty} \alpha_i^k \circ R_{i,t-k}\right) R_{j,t}\right) - \mu_{X_i}\lambda_j = \sum_{k=0}^{\infty} \alpha_i^k \mathbb{E}(R_{i,t-k}R_{j,t}) - \mu_{X_i}\lambda_j \\
&= \mathbb{E}(R_{i,t}R_{j,t}) + \sum_{k=1}^{\infty} \alpha_i^k \mathbb{E}(R_{i,t-k}R_{j,t}) - \mu_{X_i}\lambda_j \\
&= \mathbb{E}(R_{i,t}R_{j,t}) + \sum_{k=1}^{\infty} \alpha_i^k \lambda_i\lambda_j - \frac{\lambda_i\lambda_j}{1-\alpha_i} \\
&= \mathbb{C}\text{ov}(R_{i,t}, R_{j,t}).
\end{aligned}$$

Here we use the (2.3) equation, the fact that $R_{j,t-k}$ are i.i.d. in t and the property of $\mathbb{E}X_{j,t}$ from part (a).

(e) We have that

$$\begin{aligned}\mathbb{C}\mathrm{ov}(X_{j,t}, X_{j,t+h}) &= \mathbb{C}\mathrm{ov}(X_{j,t}, \alpha_j^h \circ X_{j,t} + \sum_{k=0}^{h-1} \alpha_j^k \circ R_{j,t+h-k}) \\ &= \mathbb{C}\mathrm{ov}(X_{j,t}, \alpha_j^h \circ X_{j,t}) + \mathbb{C}\mathrm{ov}(X_{j,t}, \sum_{k=0}^{h-1} \alpha_j^k \circ R_{j,t+h-k}) \\ &= \alpha_j^h \mathbb{C}\mathrm{ov}(X_{j,t}, X_{j,t}) = \alpha_j^h \sigma_{X_j}^2.\end{aligned}$$

Here, using the fact that $R_{j,t}$ are i.i.d in t and $t+h-k > t$ for $k < h$, we have that $\mathbb{C}\mathrm{ov}(X_{j,t}, \sum_{k=0}^{h-1} \alpha_j^k \circ R_{j,t+h-k}) = 0$. We have also used the last property from Theorem 2.1 to get the next-to-last equality.

(f) Using covariance and variance expressions from (b) and (d) we have that

$$\mathbb{C}\mathrm{orr}(X_{j,t+h}, X_{j,t}) = \frac{\mathbb{C}\mathrm{ov}(X_{j,t+h}, X_{j,t})}{\sqrt{\mathbb{V}\mathrm{ar}(X_{j,t+h})\mathbb{V}\mathrm{ar}(X_{j,t})}} = \frac{\alpha_j^h \sigma_{X_j}^2}{\sqrt{\sigma_{X_j}^4}} = \alpha_j^h.$$

(g) Using equations (2.2), (2.3) and Theorem 2.1 we have that

$$\begin{aligned}\mathbb{C}\mathrm{ov}(X_{i,t}, X_{j,t+h}) &= \mathbb{E}(X_{i,t}X_{j,t+h}) - \mathbb{E}(X_{i,t})\mathbb{E}(X_{j,t+h}) \\ &= \mathbb{E}\left(X_{i,t}\left[\alpha_j^h \circ X_{j,t} + \sum_{l=0}^{h-1} \alpha_j^l \circ R_{j,t+h-l}\right]\right) - \mu_{X_i}\mu_{X_j} \\ &= \alpha_j^h \mathbb{E}\left(X_{i,t}X_{j,t}\right) + \sum_{l=0}^{h-1} \alpha_j^l \mathbb{E}\left(X_{i,t}R_{j,t+h-l}\right) - \mu_{X_i}\mu_{X_j} \\ &= \alpha_j^h \mathbb{E}\left(X_{i,t}X_{j,t}\right) + (1-\alpha_j^h)\mu_{X_i}\mu_{X_j} - \mu_{X_i}\mu_{X_j} \\ &= \alpha_j^h \left(\mathbb{E}\left(X_{i,t}X_{j,t}\right) - \mu_{X_i}\mu_{X_j}\right) \\ &= \alpha_j^h \mathbb{C}\mathrm{ov}(X_{i,t}X_{j,t}) \\ &= \frac{\alpha_j^h}{1-\alpha_i\alpha_j}\mathbb{C}\mathrm{ov}(R_{i,t}, R_{j,t}),\end{aligned}$$

because:

$$\begin{aligned}\sum_{l=0}^{h-1} \alpha_j^l \mathbb{E}\left(X_{i,t}R_{j,t+h-l}\right) &= \sum_{l=0}^{h-1} \alpha_j^l \mathbb{E}(X_{i,t})\mathbb{E}(R_{j,t+h-l}) = \mu_{X_i}\lambda_j\left(\sum_{l=0}^{h-1} \alpha_j^l\right) \\ &= \mu_{X_i}\lambda_j \frac{1-\alpha_j^h}{1-\alpha_j} = (1-\alpha_j^h)\mu_{X_i}\mu_{X_j},\end{aligned}$$

and

$$\begin{aligned}\mathbb{C}\mathrm{ov}(X_{i,t}X_{j,t}) &= \mathbb{C}\mathrm{ov}\left(\sum_{k=0}^{\infty} \alpha_i^k \circ R_{i,t-k}, \sum_{s=0}^{\infty} \alpha_j^s \circ R_{j,t-s}\right) = \sum_{k,s=0}^{\infty} \alpha_i^k \alpha_j^s \mathbb{C}\mathrm{ov}(R_{i,t-k}, R_{j,t-s}) \\ &= \sum_{k=0}^{\infty} \alpha_i^k \alpha_j^k \mathbb{C}\mathrm{ov}(R_{i,t-k}, R_{j,t-k}) = \left(\sum_{k=0}^{\infty} \alpha_i^k \alpha_j^k\right)\mathbb{C}\mathrm{ov}(R_{i,t}, R_{j,t}) \\ &= \frac{1}{1-\alpha_i\alpha_j}\mathbb{C}\mathrm{ov}(R_{i,t}, R_{j,t}), \quad h \geq 0,\end{aligned}$$

where $\sum_{l=0}^{h-1} \alpha_j^l$ is a geometric series and $\sum_{k=0}^{\infty} \alpha_i^k \alpha_j^k$ is an infinite geometric series.

(h) From the definition of the correlation as well as using properties (b) and (f) we have that

$$\begin{aligned}\mathbb{C}\text{orr}(X_{i,t+h}, X_{j,t}) &= \frac{\mathbb{C}\text{ov}(X_{i,t+h}, X_{j,t})}{\sqrt{\mathbb{V}\text{ar}(X_{i,t+h})\mathbb{V}\text{ar}(X_{j,t})}} \\ &= \frac{\alpha_i^h \sqrt{(1-\alpha_i^2)(1-\alpha_j^2)}}{(1-\alpha_1\alpha_2)\sqrt{(\sigma_i^2+\alpha_i\lambda_i)(\sigma_j^2+\alpha_j\lambda_j)}} \mathbb{C}\text{ov}(R_{i,t}, R_{j,t}), \quad h \geq 0.\end{aligned}$$

From the covariance and correlation of the BINAR(1) process we see that the dependence between $X_{1,t}$ and $X_{2,t}$ depends on the joint distribution of the innovations $R_{1,t}$, $R_{2,t}$. Pedeli and Karlis (2011) analysed BINAR(1) models when the innovations were joint by either a bivariate Poisson or a bivariate negative binomial distribution, where the covariance of the innovations can be easily expressed in terms of their joint distribution parameters. Karlis and Pedeli (2013) analysed two cases when the innovations of a BINAR(1) model are joint by specific copula functions. We will expand on their work, by analysing additional copulas for the BINAR(1) model innovation distribution as well as estimation methods for the distribution parameters. We will firstly provide the definition of copulas and their properties in the next section.

3 Copulas

In this section we provide the definition and main properties of bivariate copulas mainly following Genest and Nešlehová (2007), Nelsen (2006) and Trivedi and Zimmer (2007) for the continuous and discrete settings.

3.1 Copula definition and properties

Copulas are used for modelling the dependence between several random variables. The term copula[2] was first used by Sklar (1959). The main advantage of using copulas is that they allow to model the marginal distributions separately from their joint distribution.

In general, the joint cumulative distribution function of a random vector (X_1, X_2) is defined as:

$$H(x_1, x_2) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2). \tag{3.1}$$

The marginal distributions F_1 and F_2 are defined as:

$$F_i(x) = \mathbb{P}(X_i \leq x), \quad i = 1, 2. \tag{3.2}$$

Copulas are multivariate distribution functions with uniform marginal distributions on the interval [0,1].

[2]the term copula is derived from the latin word *copulare* - to connect, to join.

Definition 3.1. A 2-dimensional copula $C : [0,1]^2 \to [0,1]$ is a function with the following properties:

1. For every $u, v \in [0,1]$:

$$C(u,0) = C(0,v) = 0 \tag{3.3}$$

2. For every $u, v \in [0,1]$:

$$C(u,1) = u, \quad C(1,v) = v. \tag{3.4}$$

3. For any $u_1, u_2, v_1, v_2 \in [0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$:

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0 \tag{3.5}$$

(this is also called *the rectangle inequality*).

The theoretical foundation of copulas is given by Sklar's theorem:

Theorem 3.1. ***Sklar (1959).*** *Let H be a joint cumulative distribution function (cdf) with marginal distributions F_1, F_2. Then there exists a copula C such that $\forall x_i \in [-\infty, \infty]$, $i = 1, 2$:*

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \tag{3.6}$$

If F_i is continuous $\forall i = 1, 2$ then C is unique; otherwise C is uniquely determined only on $RanF_1 \times RanF_2$, where $RanF_i$ denotes the range of the cdf F_i. Conversely, if C is a copula and F_1, F_2 are distribution functions, then the function H, defined by equation (3.6) *is a joint cdf with marginal distributions F_1, F_2.*

If a pair of random variables (X_1, X_2) has continuous marginal cdfs $F_i(x), i = 1, 2$, then by applying the probability integral transformation one can transform them into random variables $(U_1, U_2) = (F_1(X_1), F_2(X_2))$ with uniformly distributed marginals which can then be used when modelling their dependence via a copula.

Theorem 3.2. ***Probability integral transformation.*** *Let X be a random variable with a continuous cdf F. Then $U = F(X) \sim U(0,1)$, where $U(0,1)$ is the uniform distribution on the interval* $[0,1]$.

Proof: We have that $\mathbb{P}(U \leq u) = \mathbb{P}(F(X) \leq u) = \mathbb{P}(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u$, $\forall u \in (0,1)$, where $F^{-1}(u) = inf\{x : F(x) \geq y\}$. So, $U = F(X)$ has a uniform distribution on the interval $[0,1]$.

3.2 Dependence and correlation

Since copulas are used to model the dependence between parameters it is important to understand the nature of dependence captured by copulas, its relationship to correlation as well as its interpretability.

- **Correlation coefficient** can be used to measure the association between two random variables X_1 and X_2 with finite second-order moments. It is defined as:

$$\rho_{X_1,X_2} = \rho(X_1, X_2) = \frac{\mathbb{C}\text{ov}(X_1, X_2)}{\sqrt{\mathbb{V}\text{ar}(X_1)\mathbb{V}\text{ar}(X_2)}}. \tag{3.7}$$

The correlation coefficient has the following properties:

(a) ρ_{X_1,X_2} is a measure of linear dependence;

(b) ρ_{X_1,X_2} is symmetric;

(c) $-1 \leq \rho_{X_1,X_2} \leq 1$ where $\rho_{X_1,X_2} = -1$ measures perfect negative linear dependence and $\rho_{X_1,X_2} = 1$ measures perfect positive linear dependence;

(d) ρ_{X_1,X_2} is invariant with respect to linear transformations of the variables;

(e) Zero correlation does not imply independence. For example: if $X \sim N(0,1)$ and $Y = X^2$, then $\mathbb{C}\text{ov}(X,Y) = 0$, but it is clear that X and Y are dependent.

As a result, correlation has its limitations:

(a) **Zero correlation** requires that $\mathbb{C}\text{ov}(X,Y) = 0$, however **zero dependence** requires $\mathbb{C}\text{ov}(\phi_1(X), \phi_2(Y)) = 0$ for any functions ϕ_1, ϕ_2;

(b) If a distribution does not have a second order moment, then correlation cannot be defined for it (example: Students t distribution with one degree of freedom);

(c) Correlation is not invariant under strictly increasing non-linear transformations: $\rho_{T(X),T(Y)} \neq \rho_{X,Y}$ for $T : \mathbb{R} \to \mathbb{R}$.

Given these limitations some alternative measures of dependence are considered:

- **Rank correlation** There are two established measures of correlation for measuring the dependence of two random variables X_1 and X_2 with continuous cdfs F_1, F_2:

  1. **Spearman's rho** is the linear correlation between $F_1(X_1)$ and $F_2(X_2)$:

  $$\rho_S(X_1, X_2) = \rho(F_1(X_2), F_2(X_2)). \tag{3.8}$$

  2. **Kendall's tau** is defined as the difference between the probability of concordance and the probability of discordance of the random variables:

  $$\begin{aligned}\rho_\tau(X_1, X_2) &= \mathbb{P}((X_{1,1} - X_{1,2})(X_{2,1} - X_{2,2}) > 0) - \mathbb{P}((X_{1,1} - X_{1,2})(X_{2,1} - X_{2,2}) < 0) \\ &= \mathbb{P}(concordance) - \mathbb{P}(discordance),\end{aligned} \tag{3.9}$$

  where $(X_{1,1}, X_{2,1})$ and $(X_{1,2}, X_{2,2})$ are two independent pairs of random variables from F_1 and F_2. Concordance refers to the property that large values of one random variable are associated with large values of the other random variable. Discordance refers to large values of one random variable being associated with small values of the other random variable.

Both ρ_S and ρ_τ have the following properties (more can be found in Trivedi and Zimmer (2007)):

(a) ρ_S and ρ_τ are symmetric;

(b) If the random variable pair (X_1, X_2) is independent, then $\rho_S(X_1, X_2) = 0$ and $\rho_\tau(X_1, X_2) = 0$;

(c) The copula of X_1 and X_2 is the lower bound of the Frechet-Hoeffding bounds (see below) if and only if $\rho_S(X_1, X_2) = \rho_\tau(X_1, X_2) = -1$;

(d) The copula of X_1 and X_2 is the upper bound of the Frechet-Hoeffding bounds if and only if $\rho_S(X_1, X_2) = \rho_\tau(X_1, X_2) = 1$.

3.3 Frechet-Hoeffding bounds

Let $H(x_1, x_2)$ be a bivariate joint cdf with univariate marginal cdfs F_1, F_2 and each marginal distribution can take values in the range $[0, 1]$. Then, the joint cdf is bound below and above by the Frechet-Hoeffding lower and upper bounds:

1. Lower bound:

$$W := H_L(x_1, x_2) = \max(F_1(x_1) + F_2(x_2) - 1, 0). \tag{3.10}$$

2. Upper bound:

$$M := H_U(x_1, x_2) = \min\left(F_1(x_1), F_2(x_2)\right), \tag{3.11}$$

so that:

$$W \leq H(x_1, x_2) \leq M, \quad \forall x_1, x_2. \tag{3.12}$$

The Frechet-Hoeffding bounds also applies to copulas since copulas are cumulative distribution functions:

$$\max(F_1(x_1) + F_2(x_2) - 1, 0) \leq C(F_1(x_1), F_2(x_2)) \leq \min(F_1(x_1), F_2(x_2)). \tag{3.13}$$

If the dependence parameter of a copula approaches its lower (upper) bound, then the copula should approach the Frechet-Hoeffding lower (upper) bound. However, the parametric form of a copula may impose restrictions so that one (or both) Frechet-Hoeffding bound is not included in the range.

3.4 Copulas with discrete marginal distributions

In this section we will mention some of the key differences when copula marginals are discrete rather than continuous.

Firstly, as mentioned in Theorem 3.1, if F_1 and F_2 are discrete marginals then a unique copula representation exists only for values in the range of $Ran(F_1) \times Ran(F_2)$. However, the lack of uniqueness does not pose a problem in empirical applications because it implies that there may exist more than one copula with identical properties.

Secondly, regarding concordance and discordance, the discrete case has to allow for ties, so the concordance measures are margin-dependent, see Trivedi and Zimmer (2007). There are several modifications for ρ_τ proposed, however, none of them are margin-free. Furthermore, Genest and Nešlehová (2007) states that estimations of the dependence parameter θ based on Kendall's tau or its modified versions are biased and estimation techniques based on maximum likelihood are recommended.

3.5 Some concrete copulas

In this section we will introduce bivariate copulas, which will be used in the later sections when constructing and evaluating the BINAR(1) model. We provide an explanation of the dependence structure of each copula as well as graphical examples of dependence for the discrete case.

For all the copulas discussed, the following notation is used: $u_1 := F_1(x_1)$, $u_2 := F_2(x_2)$ where F_1, F_2 are marginal cdfs of discrete random variables and θ is the dependence parameter. In the examples the discrete random variables X_1 and X_2 both follow a Poisson distribution with parameters λ_1 and λ_2 respectively:

$$F_i(k) = \mathbb{P}(X_i \leq k) = e^{-\lambda_i} \sum_{j=0}^{k} \frac{\lambda_i^j}{j!}, \quad k \in \{0, 1, 2, ...\}, \quad i = 1, 2. \tag{3.14}$$

Another discrete marginal distribution, which will be considered in the empirical application is the negative binomial distribution with an alternative parametrization in terms of its mean λ_i and variance σ_i^2:

$$F_i(k) = \mathbb{P}(X_i \leq k) = \sum_{j=0}^{k} \binom{j + \frac{\lambda_i^2}{\sigma_i^2 - \lambda_i} - 1}{j} \left(\frac{\lambda_i}{\sigma_i^2}\right)^{\left(\frac{\lambda_i^2}{\sigma_i^2 - \lambda_i}\right)} \left(\frac{\sigma_i^2 - \lambda_i}{\sigma_i^2}\right)^j, \quad i = 1, 2 \tag{3.15}$$

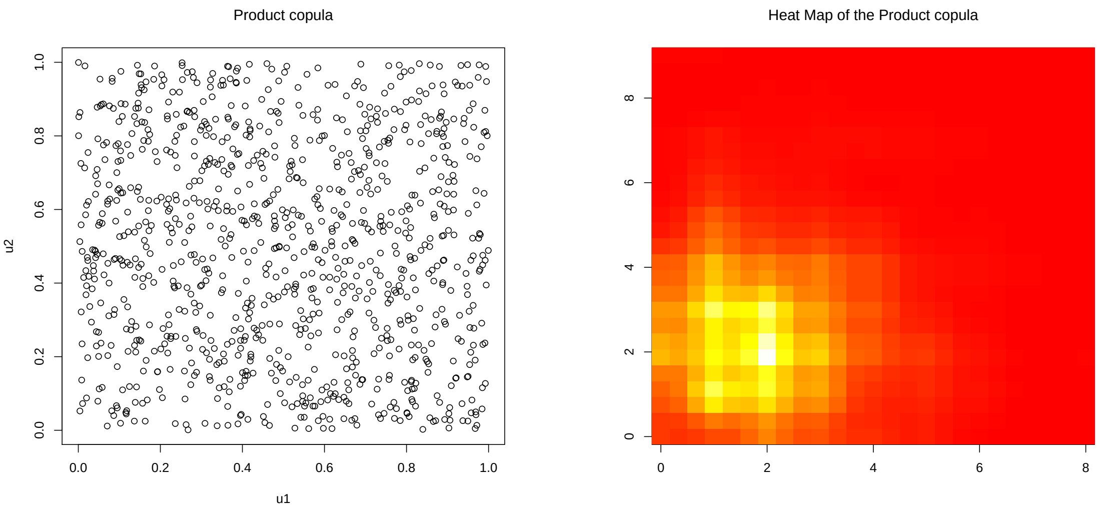
When talking about copula examples with continuous marginals, Theorem 3.2 is used, which means that $U_1 = F_1(X_1) \sim U(0,1)$ and $U_2 = F_2(X_2) \sim U(0,1)$.

3.5.1 Product copula

The product copula has the form:

$$C(u_1, u_2) = u_1 u_2. \tag{3.16}$$

The product copula corresponds to independence so it is important as a benchmark.



(a) Continuous marginals

(b) Discrete marginals

Figure 3.1: The product copula for the continuous marginal case and the discrete marginal case

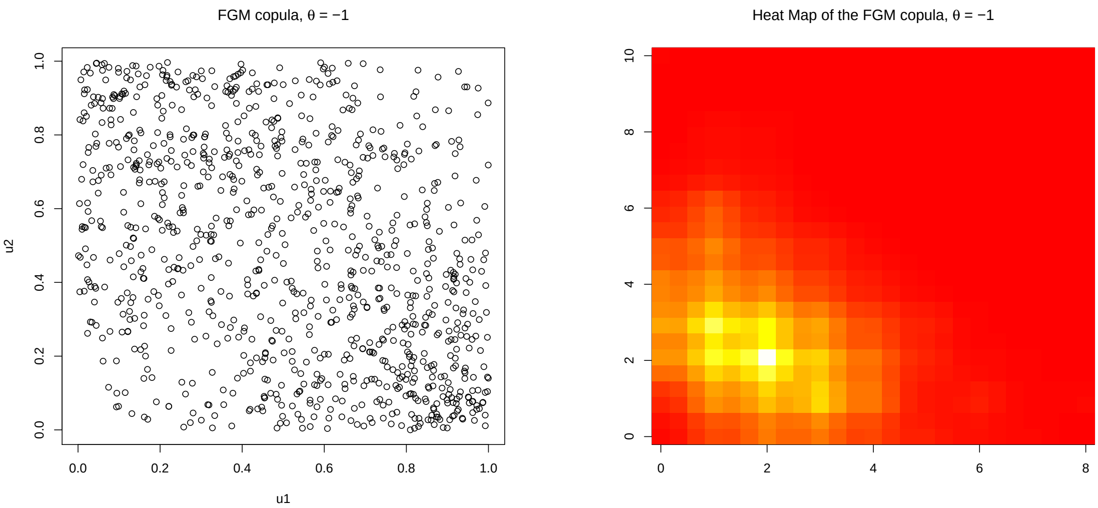
Figure A.2 provides the graphical representations of the product copula for the continuous and the discrete cases. As we can see from the plots, there does not seem to be any dependence between the two random variables.

3.5.2 Farlie-Gumbel-Morgenstern copula

The Farlie-Gumbel-Morgenstern (FGM) copula has the following form:

$$C(u_1, u_2; \theta) = u_1 u_2 (1 + \theta(1 - u_1)(1 - u_2)). \tag{3.17}$$

The dependence parameter θ can take values from the interval $[-1, 1]$. If $\theta = 0$, then the FGM copula collapses to independence. Even though the analytical form of the FGM copula is relatively simple, the FGM copula can only model weak dependence between two marginals (see Nelsen (2006)).



(a) Continuous marginals

(b) Discrete marginals

Figure 3.2: The FGM copula for the continuous marginal case and the discrete marginal case

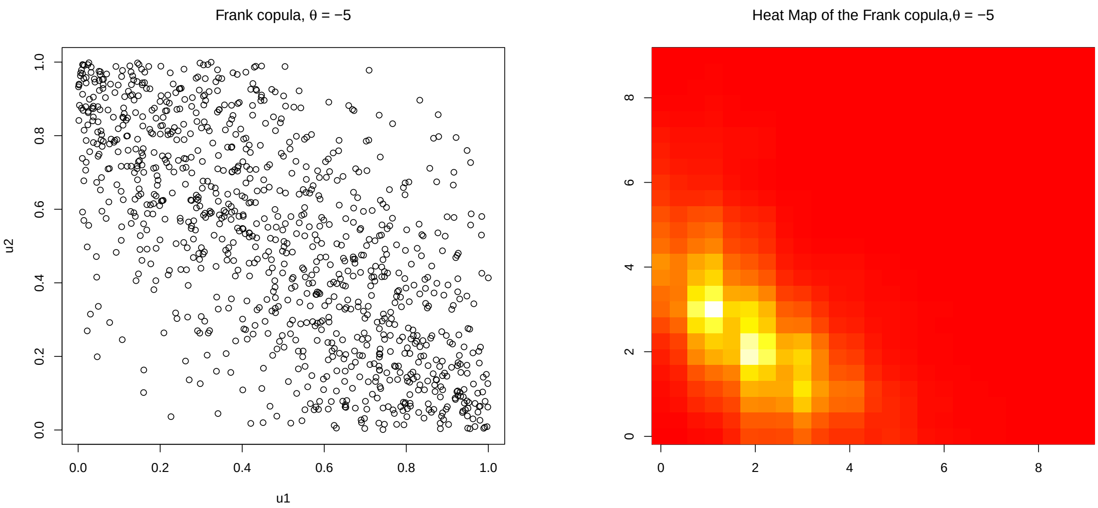
Figure 3.2 shows the FGM copula for the continuos and discrete marginal cases when the dependence parameter is -1. We can see that even when θ has the minimum value, the dependence between the random variables isn't strong.

3.5.3 Frank copula

The Frank copula has the following form:

$$C(u_1, u_2; \theta) = -\frac{1}{\theta} \log \left(1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{\exp(-\theta) - 1} \right). \tag{3.18}$$

The dependence parameter can take values from $(-\infty, \infty) \setminus \{0\}$. The Frank copula allows for both positive and negative dependence between the marginals.



(a) Continuous marginals (b) Discrete marginals

Figure 3.3: The Frank copula for the continuous marginal case and the discrete marginal case

Figure 3.3 shows the Frank copula for the continuous and discrete marginal cases when $\theta = -5$. We can see from the figures that the negative dependence is clearer compared to the FGM copula case.

3.5.4 Clayton copula

The Clayton copula has the following form:

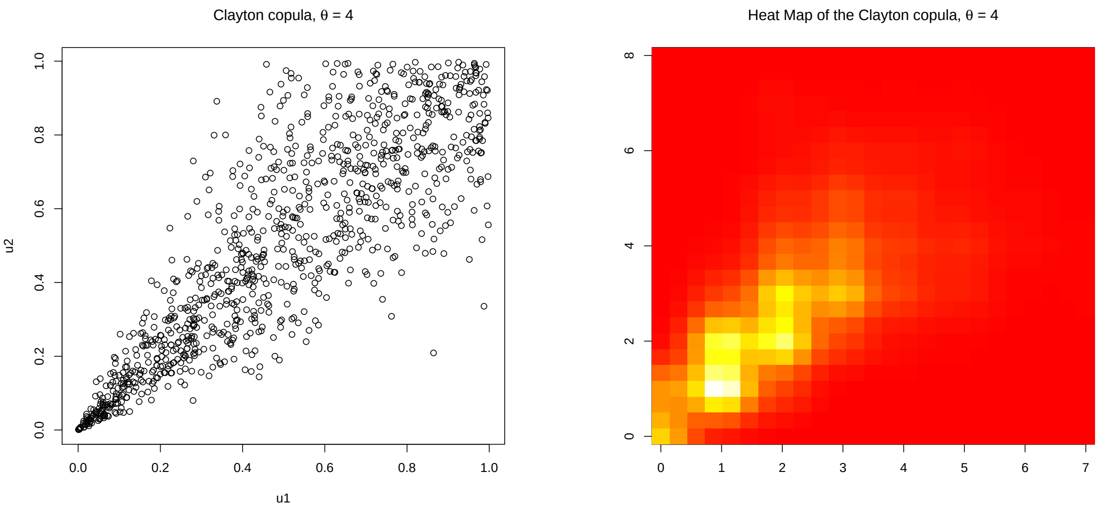
$$C(u_1, u_2; \theta) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \tag{3.19}$$

with the dependence parameter $\theta \in (0, \infty)$. The marginals become independent when $\theta \to 0$. However, the Clayton copula, defined by equation (3.19), does not account for negative dependence. It can be used when the correlation between two random variables exhibits a strong left tail dependence - if smaller values are strongly correlated and hight values are less correlated.

The Clayton copula can also be extended to account for negative dependence:

$$C(u_1, u_2; \theta) = \max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\}^{-\frac{1}{\theta}}, \tag{3.20}$$

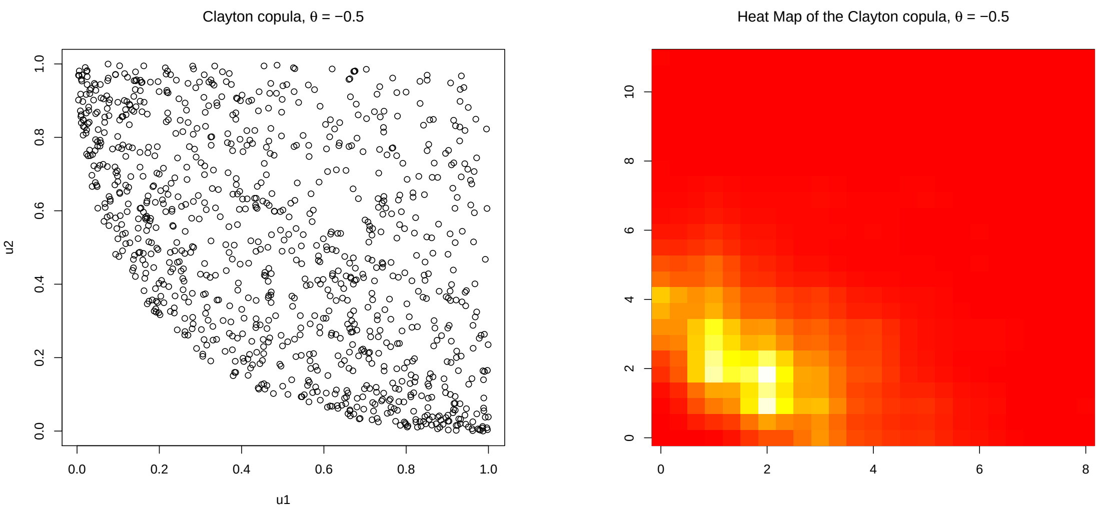
with the dependence parameter $\theta \in [-1, \infty) \setminus \{0\}$.



(a) Continuous marginals (b) Discrete marginals

Figure 3.4: The Clayton copula for the continuous marginal case and the discrete marginal case

Figure 3.4 shows the Clayton copula for the continuous and discrete marginal cases when the dependence parameter $\theta = 4$. The positive dependence between the two random variables can be seen from the plots. We can see the strong left tail dependence and the weak right tail dependence - smaller values are more correlated than large values. The case when the dependence parameter is negative ($\theta = -0.5$) is provided in Figure 3.5.



(a) Continuous marginals (b) Discrete marginals

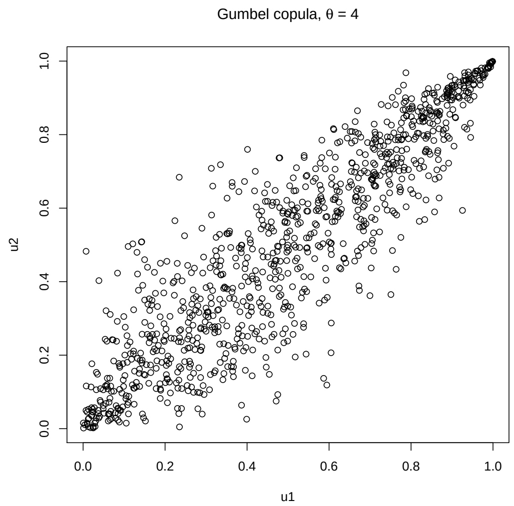
Figure 3.5: The Clayton copula for the continuous marginal case and the discrete marginal case with a negative dependence parameter

3.5.5 Gumbel copula

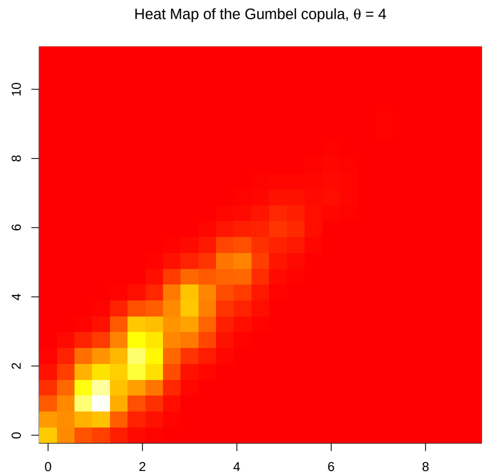
The Gumbel copula has the following form:

$$C(u_1, u_2; \theta) = \exp\left(-((-\log(u_1))^\theta + (-\log(u_2))^\theta)^{\frac{1}{\theta}}\right), \tag{3.21}$$

where the dependence parameter $\theta \in [1, \infty)$. If $\theta = 1$ then the marginals are independent. Similarly to the Clayton copula it does not allow for negative dependence. However, unlike the Clayton, the Gumbel copula exhibits strong right tail dependence and weak left tail dependence. As a result the Gumbel copula is appropriate when outcomes are strongly correlated at hight values but less correlated at low values.



(a) Continuous marginals



(b) Discrete marginals

Figure 3.6: The Gumbel copula for the continuous marginal case and the discrete marginal case

Figure 3.6 shows the Gumbel copula for the continuous and discrete marginal cases when $\theta = 4$. We can see the strong right tail dependence and the weak left tail dependence - larger values are more correlated than smaller values.

4 Model parameter estimation

In this section we examine parameter estimation methods for BINAR(1) models and for copulas. We begin by providing an estimation method for discrete copulas which is recommended in Genest and Nešlehová (2007). We also examine an additional estimation method based on minimizing the difference between the empirical joint cdf and the copula. Secondly, we analyse BINAR(1) model parameter estimation methods and provide an estimation method for estimating the copula dependence parameter separately. Estimation methods are compared via Monte Carlo simulations.

4.1 Copula parameter estimation

Following Genest and Nešlehová (2007) and Nelsen (2006), discrete copulas can be estimated via the Full Maximum Likelihood (FML) method. We also examine an estimation method based on the minimum distance between the empirical joint cdf and the theoretical copula and compare it to the FML method via Monte Carlo simulations.

4.1.1 Full Maximum likelihood estimation

Let $(x_{1,1}, x_{2,1}), ..., (x_{1,T}, x_{2,T})$ be non-negative integer valued variables and F_1, F_2 - discrete cdfs with parameters β_1, β_2 respectively and let θ be the dependence parameter of a copula. Since copulas provide the joint cdf, in order to derive the joint probability mass function (pmf), we need to take the finite differences of the copula. In the bivariate case, the joint pmf is:

$$\begin{aligned} c(F_1(x_{1,t};\beta_1), F_2(x_{2,t};\beta_2);\theta) &= \mathbb{P}(X_{1,t} = x_{1,t}, X_{2,t} = x_{2,t}) \\ &= C((F_1(x_{1,t};\beta_1), F_2(x_{2,t};\beta_2);\theta) \\ &- C((F_1(x_{1,t}-1;\beta_1), F_2(x_{2,t};\beta_2);\theta) \\ &- C((F_1(x_{1,t};\beta_1), F_2(x_{2,t}-1;\beta_2);\theta) \\ &+ C((F_1(x_{1,t}-1;\beta_1), F_2(x_{2,t}-1;\beta_2);\theta). \end{aligned} \tag{4.1}$$

Then the log-likelihood function for the FML estimation method is:

$$\mathcal{L}_T(\beta_1, \beta_2, \theta) = \sum_{t=1}^{T} \log\left(c(F_1(x_{1,t};\beta_1), F_2(x_{2,t};\beta_2);\theta)\right) \to \max_{\beta_1,\beta_2,\theta}. \tag{4.2}$$

Numerical maximizations can be carried out via the **optim** function in the R statistical software.

4.1.2 Estimation based on the distance between the empirical joint cumulative distribution function and the copula

Based on Sklar's theorem 3.1, we have that the discrete joint cdf $H(\cdot)$ can be expressed via a discrete copula $C(\cdot)$. This means that:

$$H(x_{1,t}, x_{2,t}) = C(F_1(x_{1,t};\beta_1), F_2(x_{2,t};\beta_2);\theta), \quad \forall t = 1, ..., T. \tag{4.3}$$

By substituting the joint cdf with its empirical expression we can estimate the copula parameters by minimizing the squared distance between the empirical cdf and the copula. The Empirical Distance (ED) estimator can be calculated by minimizing:

$$D(\beta_1, \beta_2, \theta) = \sum_{t=1}^{T} \left(\hat{H}(x_{1,t}, x_{2,t}) - C(F_1(x_{1,t};\beta_1), F_2(x_{2,t};\beta_2);\theta)\right)^2 \to \min_{\beta_1,\beta_2,\theta}, \tag{4.4}$$

where

$$\hat{H}(x,y) = \hat{H}_T(x,y) := \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(x_{t,1} \leq x, x_{t,2} \leq y). \tag{4.5}$$

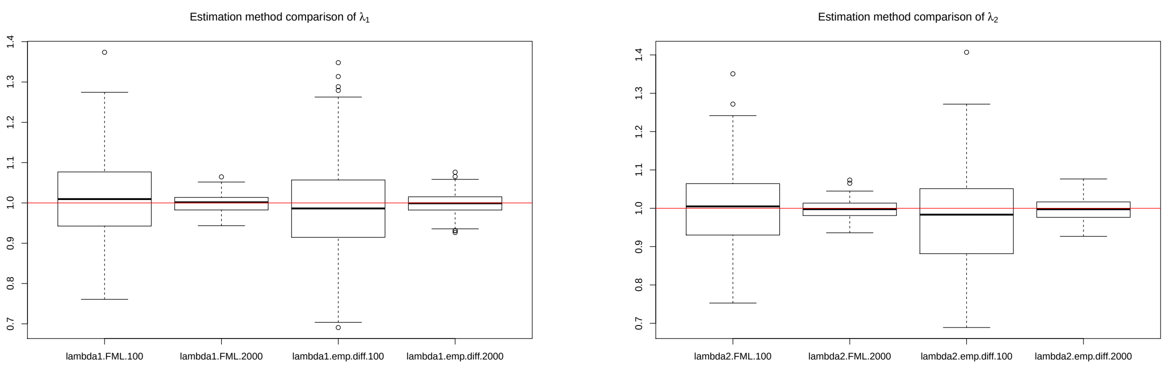
By the strong law of large numbers: $\hat{H}_T(x,y) \overset{a.s.}{\to} H(x,y), T \to \infty, \forall x, y$. Therefore, $D(\beta_1, \beta_2, \theta) \overset{a.s.}{\to} 0$ as $T \to \infty$.

4.1.3 Estimation method comparisons via Monte Carlo simulations

To illustrate the estimation methods, we run a Monte Carlo simulation 200 times for the sample size of $T = 100$ and $T = 2000$. The copula selected was an FGM copula, defined in equation (3.17) with Poisson marginals defined by eq. (3.14). The true parameter vector is $(\lambda_1, \lambda_2, \theta) = (1, 1, 0.5)$.

			FML estimator		ED estimator	
Sample size	Parameter	True Value	MSE	Bias	MSE	Bias
$T = 100$	λ_1	1	0.00968	0.01173	0.01349	-0.01109
	λ_2	1	0.01020	-0.00175	0.01553	-0.02677
	θ	0.5	0.10188	0.00364	0.39489	-0.21027
$T = 2000$	λ_1	1	0.00058	-0.00115	0.00074	-0.00032
	λ_2	1	0.00057	-0.00292	0.00091	-0.00385
	θ	0.5	0.00458	-0.00420	0.01828	-0.00488

Table 4.1: Monte Carlo simulation results



(a) Boxplots of λ_1 estimates

(b) Boxplots of λ_2 estimates

Figure 4.1: Monte Carlo simulation results of FML and ED estimators.

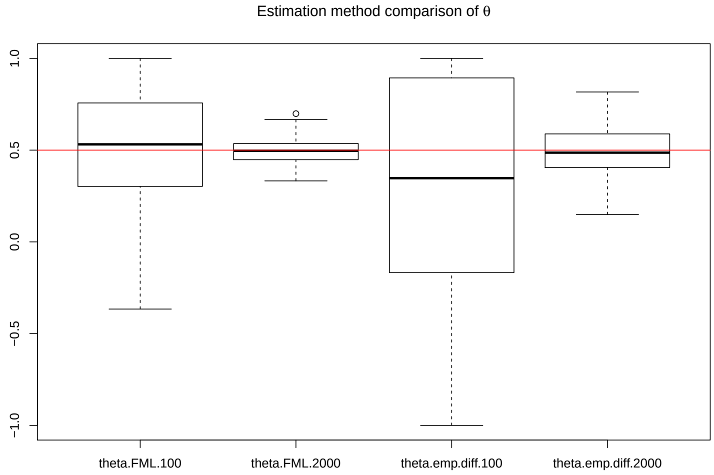


Figure 4.2: Monte Carlo simulation results of FML and ED estimators of θ.

The Monte Carlo simulations show that the FML estimates are closer to the actual parameter values when the sample size is small and the estimate boxplots from Figure 4.2 show that the ED estimates are more scattered compared to the FML estimates. However, for a larger sample size, the difference between the ED estimates and the FML estimates is smaller although the FML estimation method is still superior in terms of its smaller mean squared error and bias.

4.2 Parameter estimation of the BINAR(1) model with copula-distributed innovations

Let $\mathbf{X}_t = (X_{1,t}, X_{2,t})'$ be a non-negative integer-valued time series given in Def. 2.1 and $\mathbf{R}_t = (R_{1,t}, R_{2,t})'$ be non-negative integer-valued random variables where $R_{j,t}$ has a finite mean λ_j and variance σ_j^2, $j = 1, 2$ and let $\mathbb{C}\mathrm{ov}(R_{i,t}, R_{j,s}) = 0$, $t \neq s$, $\forall i, j \in \{1, 2\}$ with marginals F_1, F_2 and their joint cdf linked by a copula $C(\cdot, \cdot)$.

4.2.1 Conditional least squares estimation (CLS)

The Conditional Least Squares estimation minimizes the squared distance between $\mathbf{X}_t$ and its conditional expectation. We will follow the method provided in Silva (2005) for the INAR(1) model and show that the parameter estimation does not differ for the BINAR(1) model.

Using Theorem 2.1 we can define the vector of conditional means, conditionally on the previous observations:

$$\boldsymbol{\mu}_{t|t-1} := \begin{bmatrix} \mathbb{E}(X_{1,t}|X_{1,t-1}) \\ \mathbb{E}(X_{2,t}|X_{2,t-1}) \end{bmatrix} = \begin{bmatrix} \alpha_1 X_{1,t-1} + \lambda_1 \\ \alpha_2 X_{2,t-1} + \lambda_2 \end{bmatrix}. \tag{4.6}$$

We also note that:

$$\begin{aligned}
\mathbb{E}(X_{j,t}|X_{i,t-1},X_{j,t-1}) &= \mathbb{E}(\alpha_j \circ X_{j,t-1} + R_{j,t}|X_{i,t-1},X_{j,t-1}) = \mathbb{E}(\alpha_j \circ X_{j,t-1}|X_{i,t-1},X_{j,t-1}) + \lambda_j \\
&= \mathbb{E}\left(\sum_{k=1}^{X_{j,t-1}} Y_k|X_{i,t-1},X_{j,t-2}\right) + \lambda_j = \mathbb{E}\left(\sum_{k=1}^{\infty} Y_k \mathbf{1}(k \le X_{j,t-1})|X_{i,t-1},X_{j,t-2}\right) + \lambda_j \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left(Y_k \mathbf{1}(k \le X_{j,t-1})|X_{i,t-1},X_{j,t-2}\right) + \lambda_j = \sum_{k=1}^{\infty} \mathbf{1}(k \le X_{j,t-1})\mathbb{E}\left(Y_k|X_{i,t-1},X_{j,t-2}\right) + \lambda_j \\
&= \alpha_j X_{j,t-1} + \lambda_j = \mathbb{E}(X_{j,t}|X_{j,t-1}).
\end{aligned}$$

In order to calculate the CLS estimators of $(\alpha_1, \alpha_2, \lambda_1, \lambda_2)'$, we define the vector of residuals as the difference between the observations and their conditional expectation:

$$\mathbf{X}_t - \boldsymbol{\mu}_{t|t-1} = \begin{bmatrix} X_{1,t} - \alpha_1 X_{1,t-1} - \lambda_1 \\ X_{2,t} - \alpha_2 X_{2,t-1} - \lambda_2 \end{bmatrix}.$$

Following Silva (2005), CLS estimators of a BINAR(1) model are found by minimizing the residuals:

$$q_j(\alpha_1, \lambda_1) := \sum_{t=2}^{N} (X_{j,t} - \alpha_j X_{j,t-1} - \lambda_j)^2 \to \min_{\alpha_j,\lambda_j}, \quad j = 1, 2. \tag{4.7}$$

Taking the derivatives with respect to α_j and λ_j, $j = 1, 2$ and equating them to zero we have:

$$\begin{aligned}
\frac{\partial q_j}{\partial \lambda_j} &= \sum_{t=2}^{N} -2(X_{j,t} - \lambda_j X_{j,t-1} - \lambda_j) = 0, \\
\lambda_j &= \frac{1}{N-1} \sum_{t=2}^{N} (X_{j,t} - \alpha_j X_{j,t-1}) = \frac{1}{N-1}\left(\sum_{t=2}^{N} X_{j,t} - \alpha_j \sum_{t=2}^{N} X_{j,t-1}\right), \qquad (4.8) \\
\frac{\partial q_j}{\partial \alpha_j} &= \sum_{t=2}^{N} -2X_{j,t-1}(X_{j,t} - \lambda_j X_{j,t-1} - \lambda_j) = 0, \\
\sum_{t=2}^{N} & X_{j,t}X_{j,t-1} - \alpha_j \sum_{t=2}^{N} X_{j,t-1}^2 - \lambda_j \sum_{t=2}^{N} X_{j,t-1} = 0. \qquad (4.9)
\end{aligned}$$

By substituting λ_j expression from (4.8) into equation (4.9) we get:

$$\begin{aligned}
&\sum_{t=2}^{N} X_{j,t}X_{j,t-1} - \alpha_j\left(\sum_{t=2}^{N} X_{j,t-1}^2 - \frac{1}{N-1}\sum_{t=2}^{N} X_{j,t-1}\sum_{t=2}^{N} X_{j,t-1}\right) - \frac{1}{N-1}\sum_{t=2}^{N} X_{j,t}\sum_{t=2}^{N} X_{j,t-1} = 0, \\
&\alpha_j\left(\sum_{t=2}^{N} X_{j,t-1}^2 - \frac{1}{N-1}\sum_{t=2}^{N} X_{j,t-1}\sum_{t=2}^{N} X_{j,t-1}\right) = \sum_{t=2}^{N} X_{j,t}X_{j,t-1} - \frac{1}{N-1}\sum_{t=2}^{N} X_{j,t}\sum_{t=2}^{N} X_{j,t-1}, \\
&\alpha_j\left(\sum_{t=2}^{N} X_{j,t-1}^2 - (N-1)(\bar{X}_j)^2\right) = \sum_{t=2}^{N}\left(X_{j,t}X_{j,t-1} - X_{j,t}\bar{X}_j\right), \quad \bar{X}_j = \frac{1}{N-1}\sum_{t=2}^{N} X_{j,t-1}, \\
&\alpha_j \sum_{t=2}^{N} (X_{j,t-1} - \bar{X}_j)^2 = \sum_{t=2}^{N} (X_{j,t} - \bar{X}_j)(X_{j,t-1} - \bar{X}_j).
\end{aligned}$$

Finally we get:

$$\hat{\alpha}_j^{CLS} = \frac{\sum_{t=2}^{N}(X_{j,t} - \bar{X}_j)(X_{j,t-1} - \bar{X}_j)}{\sum_{t=2}^{N}(X_{j,t-1} - \bar{X}_j)^2}, \tag{4.10}$$

and substituting $\hat{\alpha}_j^{CLS}$ into (4.8):

$$\hat{\lambda}_j^{CLS} = \frac{1}{N-1}\left(\sum_{t=2}^{N} X_{j,t} - \hat{\alpha}_j^{CLS}\sum_{t=2}^{N} X_{j,t-1}\right). \tag{4.11}$$

We have that the CLS estimators of α_j and λ_j do not depend on the innovation dependence parameter θ. In order to estimate θ we notice that:

$$\begin{aligned}
&\mathbb{E}\left((X_{1,t} - \alpha_1 X_{1,t-1} - \lambda_1)(X_{2,t} - \alpha_2 X_{2,t-1} - \lambda_2)\right) \\
&\quad = \mathbb{E}\left((X_{1,t} - \alpha_1 X_{1,t-1} - \lambda_1)(X_{2,t} - \alpha_2 X_{2,t-1} - \lambda_2)\right) \\
&\quad = \mathbb{E}\left[((\alpha_1 \circ X_{1,t-1}) - \alpha_1 X_{1,t-1} + (R_{1,t} - \lambda_1))((\alpha_2 \circ X_{2,t-1} - \alpha_2 X_{2,t-1}) + (R_{2,t} - \lambda_2)\right] \\
&\quad = \mathbb{E}\left[(\alpha_1 \circ X_{1,t-1} - \alpha_1 X_{1,t-1})(\alpha_2 \circ X_{2,t-1} - \alpha_2 X_{2,t-1})\right] \\
&\quad + \mathbb{E}\left[(\alpha_1 \circ X_{1,t-1} - \alpha_1 X_{1,t-1})(R_{2,t} - \lambda_2)\right] + \mathbb{E}\left[(\alpha_2 \circ X_{2,t-1} - \alpha_2 X_{2,t-1})(R_{1,t} - \lambda_1)\right] \\
&\quad + \mathbb{E}\left[(R_{1,t} - \lambda_1)(R_{2,t} - \lambda_2)\right] = \mathbb{C}\mathrm{ov}(R_{1,t}, R_{2,t}),
\end{aligned} \tag{4.12}$$

because :

$$\begin{aligned}
&\mathbb{E}\left[(\alpha_1 \circ X_{1,t-1} - \alpha_1 X_{1,t-1})(\alpha_2 \circ X_{2,t-1} - \alpha_2 X_{2,t-1})\right] \\
&\quad = \mathbb{E}\left[(\alpha_1 \circ X_{1,t-1})(\alpha_2 \circ X_{2,t-1})\right] - \alpha_2 \mathbb{E}\left[(\alpha_1 \circ X_{1,t-1})X_{2,t-1}\right] - \alpha_1 \mathbb{E}\left[(\alpha_2 \circ X_{2,t-1})X_{1,t-1}\right] \\
&\quad + \alpha_1\alpha_2 \mathbb{E}\left[X_{1,t-1}X_{2,t-1}\right] = \alpha_1\alpha_2\mathbb{E}(X_{1,t-1}X_{2,t-1}) - \alpha_1\alpha_2\mathbb{E}(X_{1,t-1}X_{2,t-1}) \\
&\quad - \alpha_1\alpha_2\mathbb{E}(X_{1,t-1}X_{2,t-1}) + \alpha_1\alpha_2\mathbb{E}(X_{1,t-1}X_{2,t-1}) = 0.
\end{aligned}$$

From properties (f) and (h) in Theorem 2.1 we have:

$$\begin{aligned}
&\mathbb{E}\left[(\alpha_1 \circ X_{1,t-1} - \alpha_1 X_{1,t-1})(R_{2,t} - \lambda_2)\right] \\
&\quad = \mathbb{E}\left[(\alpha_1 \circ X_{1,t-1})R_{2,t}\right] - \mathbb{E}\left[(\alpha_1 \circ X_{1,t-1})\lambda_2\right] - \mathbb{E}\left[\alpha_1 X_{1,t-1}R_{2,t}\right] + \mathbb{E}\left[\alpha_1 X_{1,t-1}\lambda_2\right] \\
&\quad = \alpha_1\mathbb{E}X_{1,t-1}\mathbb{E}R_{2,t} - \alpha_1\lambda_2\mathbb{E}X_{1,t-1} - \alpha_1\mathbb{E}X_{1,t-1}\mathbb{E}R_{2,t} + \alpha_1\lambda_2\mathbb{E}X_{1,t-1} = 0, \\
&\mathbb{E}\left[(\alpha_2 \circ X_{2,t-1} - \alpha_2 X_{2,t-1})(R_{1,t} - \lambda_1)\right] \\
&\quad = \mathbb{E}\left[(\alpha_2 \circ X_{2,t-1})R_{1,t}\right] - \mathbb{E}\left[(\alpha_2 \circ X_{2,t-1})\lambda_1\right] - \mathbb{E}\left[\alpha_2 X_{2,t-1}R_{1,t}\right] + \mathbb{E}\left[\alpha_2 X_{2,t-1}\lambda_1\right] \\
&\quad = \alpha_2\mathbb{E}X_{2,t-1}\mathbb{E}R_{1,t} - \alpha_2\lambda_1\mathbb{E}X_{2,t-1} - \alpha_2\mathbb{E}X_{2,t-1}\mathbb{E}R_{1,t} + \alpha_2\lambda_1\mathbb{E}X_{2,t-1} = 0, \\
&\mathbb{E}\left[(R_{1,t} - \lambda_1)(R_{2,t} - \lambda_2)\right] = \mathbb{E}\left[R_{1,t}R_{2,t}\right] - \mathbb{E}\left[R_{1,t}\lambda_2\right] - \mathbb{E}\left[R_{2,t}\lambda_1\right] + \lambda_1\lambda_2 \\
&\quad = \mathbb{E}\left[R_{1,t}R_{2,t}\right] - \lambda_1\lambda_2 - \lambda_1\lambda_2 + \lambda_1\lambda_2 \\
&\quad = \mathbb{E}\left[R_{1,t}R_{2,t}\right] - \lambda_1\lambda_2 = \mathbb{C}\mathrm{ov}(R_{1,t}, R_{2,t}).
\end{aligned}$$

For the Poisson marginal case, the innovations $R_{1,t}$ and $R_{2,t}$ are joint by a copula with the dependence parameter θ, so we can estimate the dependence parameter by minimizing the squared difference:

$$S = \sum_{t=2}^{N}\left((X_{1,t} - \hat{\alpha}_1^{CLS}X_{1,t-1} - \hat{\lambda}_1^{CLS})(X_{2,t} - \hat{\alpha}_2^{CLS}X_{2,t-1} - \hat{\lambda}_2^{CLS}) - \mathbb{C}\mathrm{ov}(R_{1,t}, R_{2,t})\right)^2 \to \min_{\theta} \tag{4.13}$$

From the definition of covariance and equation (4.1) we have:

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(R_{1,t}, R_{2,t}) &= \mathbb{E}(R_{1,t}R_{2,t}) - \mathbb{E}R_{1,t}\mathbb{E}R_{2,t} \\
&= \sum_{k=0}^{\infty}\sum_{s=0}^{\infty} k \cdot s \cdot c(F_1(k), F_2(s); \theta) - \lambda_1\lambda_2.
\end{aligned}$$

We can approximate the covariance:

$$\mathbb{C}\text{ov}(R_{1,t}, R_{2,t}) \approx \sum_{k=0}^{M}\sum_{s=0}^{M} k \cdot s \cdot c(F_1(k; \hat{\lambda}_1^{CLS}), F_2(s; \hat{\lambda}_2^{CLS}); \theta) - \hat{\lambda}_1^{CLS}\hat{\lambda}_2^{CLS}. \tag{4.14}$$

Depending on the values $\hat{\lambda}_1^{CLS}$ and $\hat{\lambda}_2^{CLS}$, for values $k, s > M$: $\mathbb{P}(R_{1,t} = k, R_{2,t} = s) = 0$. Different values can be selected for k and s such that $\mathbb{P}(R_{1,t} = k, R_{2,t} = s) = 0$ for $k > M_1$ and $s > M_2$. In Figure 4.3 we can see that if the marginals are Poisson with parameters equal to 1, then the covariance of innovations joint with FGM copula is approximated after setting $M_1 = M_2 = M \approx 8$, regardless of the dependence parameter θ.

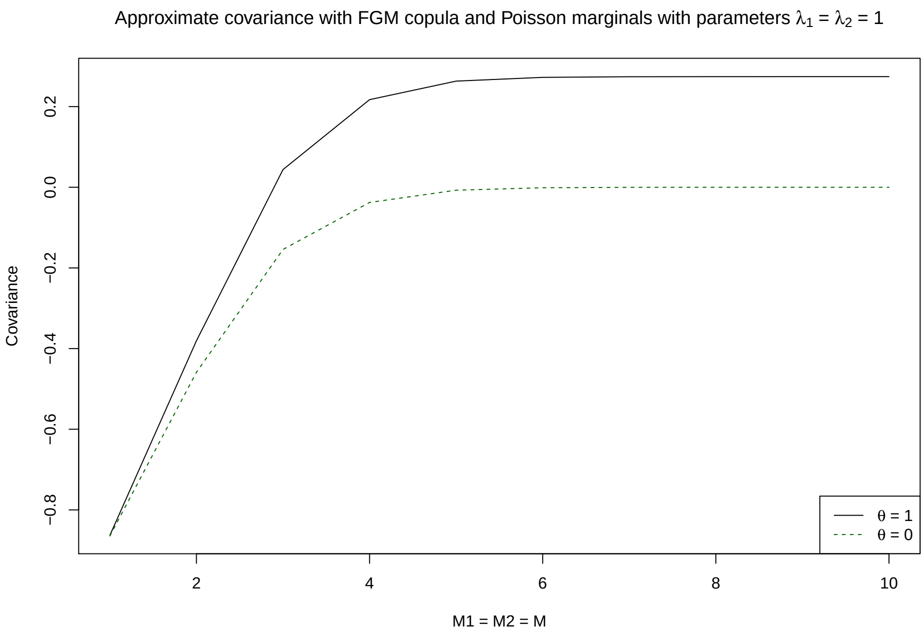


Figure 4.3: Approximate covariance with Poisson marginals and FGM copula, $\lambda_1 = \lambda_2 = 1$

Depending on the selected copula family, calculating (4.1) to get the analytical expression on $\hat{\theta}^{CLS}$ may be difficult. However, we can use the **optim** package from R to minimize (4.13). For other marginal distribution cases, where the marginal distribution has parameters other than λ_j, equation (4.13) would need to be minimized by those additional parameters. Let the additional parameter vector of the marginal cdf of innovation $R_{j,t}$ be β_j. Then:

$$S \to \min_{\beta_1, \beta_2, \theta} \tag{4.15}$$

The asymptotic properties of the CLS estimators for the INAR(1) model case are provided in Silva (2005).

4.2.2 Conditional maximum likelihood estimation (CML)

Following Pedeli and Karlis (2011) and Karlis and Pedeli (2013), BINAR(1) models can be estimated via conditional maximum likelihood (CML). The conditional distribution of the BINAR(1) process is:

$$\begin{aligned}
&\mathbb{P}(X_{1,t} = x_{1,t}, X_{2,t} = x_{2,t} | X_{1,t-1} = x_{1,t-1}, X_{2,t-1} = x_{2,t-1}) \\
&\quad = \mathbb{P}(\alpha_1 \circ X_{1,t-1} + R_{1,t} = x_{1,t}, \alpha_2 \circ X_{2,t-1} + R_{2,t} = x_{2,t} | X_{1,t-1} = x_{1,t-1}, X_{2,t-1} = x_{2,t-1}) \\
&\quad = \mathbb{P}(\alpha_1 \circ x_{1,t-1} + R_{1,t} = x_{1,t}, \alpha_2 \circ x_{2,t-1} + R_{2,t} = x_{1,t}) \qquad (4.16)
\end{aligned}$$

Let us calculate:

$$\mathbb{P}(X_{1,t} = x_{1,t}, X_{2,t} = x_{2,t} | X_{1,t-1} = x_{1,t-1}, X_{2,t-1} = x_{2,t-1}) = f(X_{1,t-1}, X_{2,t-1}),$$

where

$$\begin{aligned} f(x, y) &= \mathbb{P}(\alpha_1 \circ x + R_{1,t} = x_{1,t}, \alpha_2 \circ y + R_{2,t} = x_{2,t}) \\ &= \sum_{k=0}^{x_{1,t}} \sum_{l=0}^{x_{2,t}} \mathbb{P}(\alpha_1 \circ x + R_{1,t} = x_{1,t}, \alpha_2 \circ y + R_{2,t} = x_{2,t}, \alpha_1 \circ x = k, \alpha_2 \circ y = l) \\ &= \sum_{k=0}^{x_{1,t}} \sum_{l=0}^{x_{2,t}} \mathbb{P}(\alpha_1 \circ x = k)\mathbb{P}(\alpha_2 \circ y = l)\mathbb{P}(R_{1,t} = x_{1,t} - k, R_{2,t} = x_{2,t} - l) \end{aligned}$$

since $\alpha_j \circ x$ is the sum of x independent Bernoulli trials - it has the binomial distribution:

$$\mathbb{P}(\alpha_j \circ x = k) = \mathbb{P}\left(\sum_{i=1}^{x} Y_i = k\right) = \begin{cases} 0, & \text{if } x < k \\ \binom{x}{k}\alpha_j^k (1-\alpha_j)^{x-k}, & \text{if } x \geq k \end{cases} \tag{4.17}$$

and

$$\mathbb{P}(R_{1,t} = x_{1,t} - k, R_{2,t} = x_{2,t} - l) = c(x_{1,t} - k, x_{2,t} - s), \tag{4.18}$$

where $c(x_{1,t} - k, x_{2,t} - s)$ is the copula pmf, defined in eq. (4.1). We have that

$$\begin{aligned} &\mathbb{P}(X_{1,t} = x_{1,t}, X_{2,t} = x_{2,t} | X_{1,t-1} = x_{1,t-1}, X_{2,t-1} = x_{2,t-1}) \\ &\quad = \sum_{k=0}^{x_{1,t}} \sum_{l=0}^{x_{2,t}} \mathbb{P}(\alpha_1 \circ x_{1,t-1} = k)\mathbb{P}(\alpha_2 \circ x_{2,t-1} = l)\mathbb{P}(R_{1,t} = x_{1,t} - k, R_{2,t} = x_{2,t} - l). \end{aligned} \tag{4.19}$$

Then, the log conditional likelihood function can be defined as:

$$l(\alpha_1, \alpha_2, \lambda_1, \lambda_2, \theta) = \sum_{t=2}^{N} \log\left(\mathbb{P}(X_{1,t} = x_{1,t}, X_{2,t} = x_{2,t} | X_{1,t-1} = x_{1,t-1}, X_{2,t-1} = x_{2,t-1})\right) \tag{4.20}$$

for some initial values $x_{1,1}$ and $x_{2,1}$.

Since the likelihood depends on the marginal distribution parameters λ_1, λ_2, the probabilities of the Bernoulli trial successes α_1, α_2 and the dependence parameter θ, then in order to estimate the parameters we need to maximize the log conditional likelihood:

$$l(\alpha_1, \alpha_2, \lambda_1, \lambda_2, \theta) \to \max_{\alpha_1, \alpha_2, \lambda_1, \lambda_2, \theta}. \tag{4.21}$$

Numerical maximization is straightforward with the **optim** function from R statistical software.

As with the CLS estimator case, for other marginal distribution cases where the marginal distribution has parameters other than λ_j, equation (4.21) would need to be minimized by those additional parameters. Let the additional parameter vector of the marginal cdf of innovation $R_{j,t}$ be β_j. Then:

$$l(\alpha_1, \alpha_2, \lambda_1, \lambda_2, \beta_1, \beta_2, \theta) \to \max_{\alpha_1, \alpha_2, \lambda_1, \lambda_2, \beta_1, \beta_2, \theta}. \tag{4.22}$$

Asymptotic properties of the CML estimator are provided in Pedeli and Karlis (2011).

4.2.3 Two-stage estimation based on CLS and CML

Depending on the range of the parameter attainable values and the sample size, CML maximization might take some time to compute. Since CLS estimators of α_j and λ_j are easily derived (compared to the CLS estimator of θ, which depends on the copula pmf form and needs to be numerically maximized), we can substitute the parameters of the marginal distributions in eq. (4.21) with CLS estimates from equations (4.10) and (4.11). Then we will only need to maximize the expression with regard to a single parameter θ. The 2-Stage estimation method steps to estimate all the parameters are as follows:

1. Estimate λ_j and α_j, $j = 1, 2$ via CLS.
2. Substitute $\hat{\lambda}_1^{CLS}, \hat{\lambda}_2^{CLS}$, $\hat{\alpha}_1^{CLS}, \hat{\alpha}_2^{CLS}$ in equation (4.21) and estimate θ via CML.

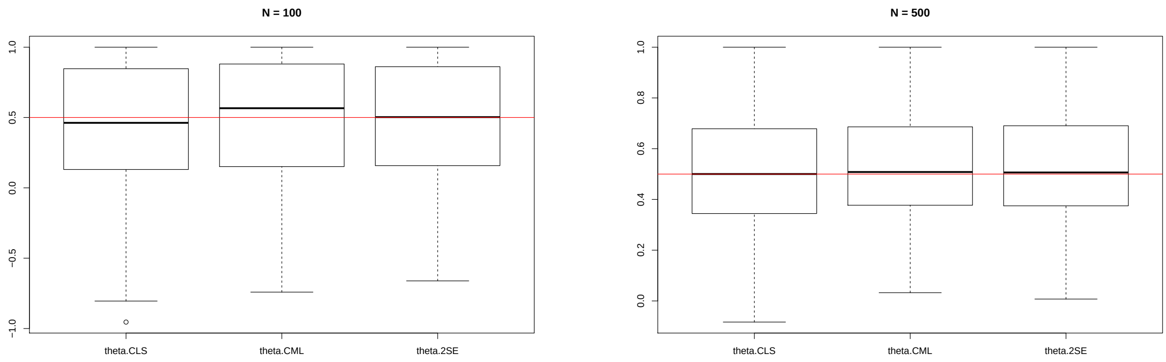
4.2.4 Estimation method comparison via Monte Carlo simulation

We carried out a Monte Carlo simulation 200 times to test the estimation methods with sample size 100 and 500. The generated model was a BINAR(1) with innovations joint by an FGM copula with Poisson marginal distributions and the parameter vector $(\alpha_1, \alpha_2, \lambda_1, \lambda_2, \theta) = (0.5, 0.8, 1, 2, 0.5)$. The results are provided in Table 4.2.

Sample size	Parameter	True value	CLS estimator		CML estimator		2-Stage estimator	
			MSE	Bias	MSE	Bias	MSE	Bias
	α_1	0.5	0.00875	-0.01932	0.00511	-0.00366	0.00875	-0.01932
	α_2	0.8	0.00426	-0.02372	0.00099	-0.00326	0.00426	-0.02372
$N = 100$	λ_1	1	0.03905	0.02697	0.02383	-0.00465	0.03905	0.02697
	λ_2	2	0.40501	0.21427	0.08672	0.01797	0.40501	0.21427
	θ	0.5	0.21944	-0.07898	0.04549	-0.02335	0.04611	-0.03345
	α_1	0.5	0.00146	-0.00114	0.00080	-0.00042	0.00146	-0.00114
	α_2	0.8	0.00077	-0.00483	0.00022	-0.00115	0.00077	-0.00483
$N = 500$	λ_1	1	0.00713	0.00427	0.00409	0.00262	0.00713	0.00427
	λ_2	2	0.07703	0.05385	0.01955	0.01594	0.07703	0.05385
	θ	0.5	0.05466	0.02040	0.04549	0.02792	0.04611	0.02749

Table 4.2: Monte Carlo simulation results

It is worth noting that CML estimation via numerical maximization depends heavily on the starting parameter values. If the starting values are selected too low, then the global maximum is not found. In order to overcome this, we have selected the starting values equal to the CLS parameter estimates. As a result, the CML estimates have a lower MSE and bias compared to the CLS estimates. However, the estimates of θ via the 2-Stage estimation method are very close to the CML estimates in terms of MSE and bias. Furthermore, since in 2-Stage estimation numerical maximization is only carried out via a single parameter θ, the starting parameter values have less of an effect on the numerical maximization. For the 2-Stage estimation method the estimates of α_j and λ_j are calculated via the CLS method and used in estimating the parameter θ via CML.



(a) Boxplots of θ estimate, $N = 100$

(b) Boxplots of θ estimate, $N = 500$

Figure 4.4: Monte Carlo simulation results.

Figure 4.4 shows the boxplots of θ estimates via different estimation methods. The rest of the parameter boxplots can be found in Appendix A. Generally, all estimators perform quite well: the median of the estimates is close to the true parameter values (which are represented by horizontal lines). Compared to CLS estimates, the interquartile range is smaller for the CML estimates of α_j and λ_j parameters. We see that for both sample sizes the 2-Stage estimator of the dependence parameter θ is very close to the CML estimator in terms of the interquartile range. In fact, the interquartile range of the dependence parameter estimate is very similar for all different estimation methods.

We can conclude that, while estimations of α_j and λ_j are closer to the true parameter values via CML estimation method, it is possible to use other estimation methods to estimate the dependence parameter with a small loss of accuracy.

5 Application of default loan data

In this section we estimate a BINAR(1) model with the joint innovation distribution modelled by a copula cdf for empirical data. The dataset consists of daily loan data which includes loans that have defaulted (i.e. they have missed 3 consecutive loan monthly payments) and loans that haven't defaulted. We will analyse and model the dependence between daily loan defaults and non-defaulted loans as well as the presence of autocorrelation.

5.1 Loan default data

The data sample used is from an Estonian peer-to-peer lending company, Bondora. In November of 2014 Bondora introduced a loan rating system which assigns a loan to a different group based on its risk level. There are a total of 8 groups ranging from the lowest risk - "AA" group - to the highest risk - "HR" group. However, the loan rating system could not be applied to all older loans due to a lack of data needed for Bondoras rating model[3]. Since a new rating model indicates new rules for accepting or rejecting loans, we have selected the data sample from 2014-04-15 because from that date there were only a few defaulted loans that did not have a rating to 2015-02-15 because of the fact that a default is considered when there are 3 consecutive missed payments. Any latter dates, which are closer to the date that this thesis was written, would have fewer loan defaults because some defaulted loans have

[3]see: https://www.bondora.com/blog/explaining-bondora-rating/, section *"What will happen to the statistics pages and data export? Will the old loans be re-evaluated?"*.

had payments made in the first few months. We are analysing data consisting of 307 daily observations which include:

- The amount of loans that were issued each day which have not defaulted (variable 'CompletedLoans');
- The amount of loans that were issued each day which have defaulted (variable 'DefaultedLoans').

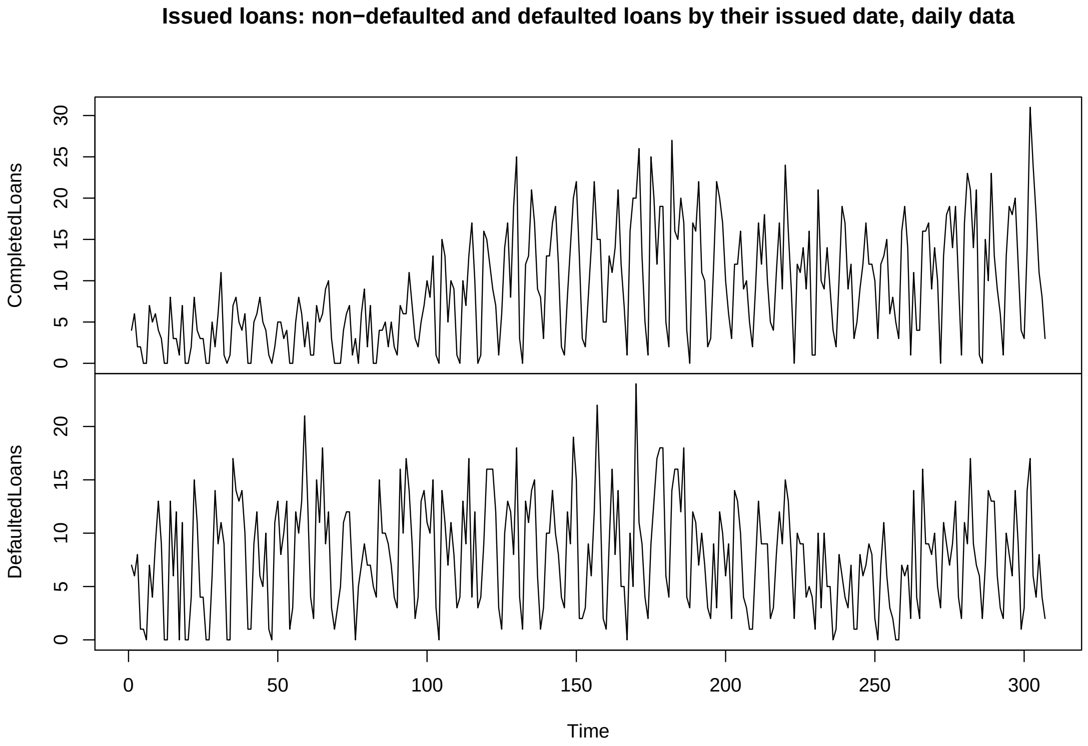


Figure 5.1: Bondora loan data: non-defaulted and defaulted loans by their issued date.

We note that the first 100 days, the mean and variance of non-defaulted loans was lower compared to the mean and variance of later days. The change could be due to a variety of reasons: the effect of the new loan rating system, which was officially implemented in December of 2014, the effect of advertising or the fact that the amount of loans, issued to people living outside of Estonia, increased. The analysis of the significance of these effects is left for future research.

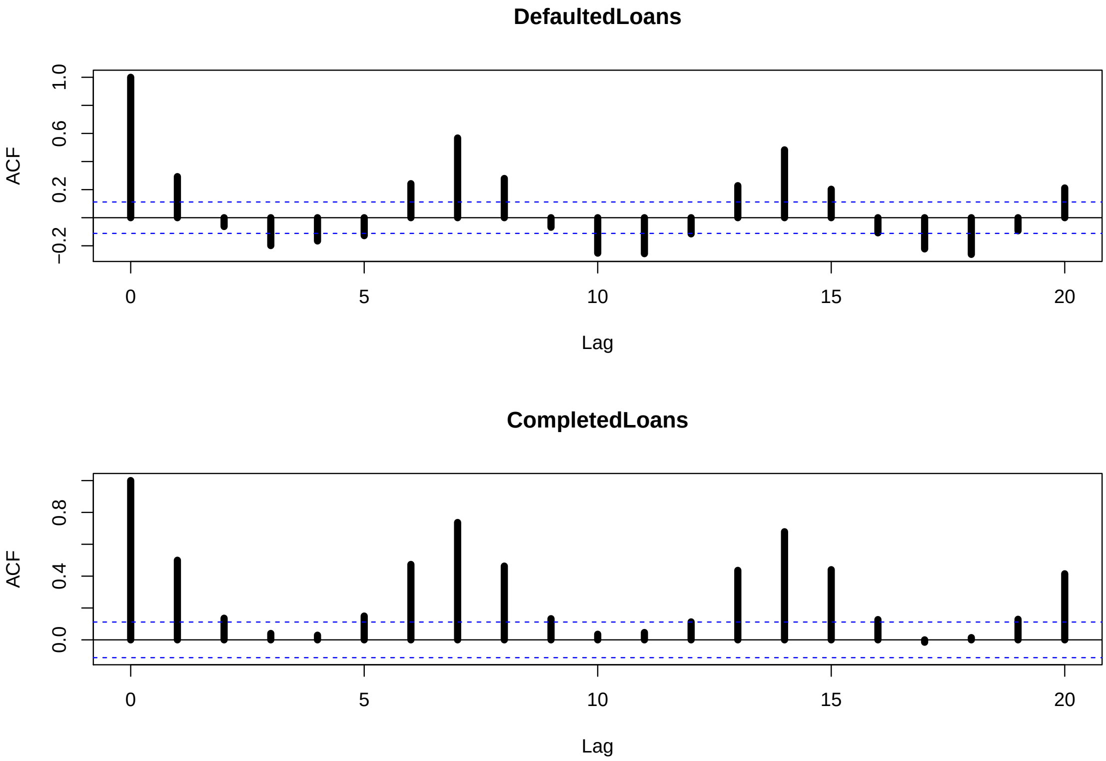


Figure 5.2: Sample autocorrelation function plots.

	min	max	mean	variance
DefaultedLoans	0.00	24.00	7.89	25.77
CompletedLoans	0.00	31.00	8.92	46.29

Table 5.1: Summary statistics of the daily data of defaulted loans and non-defaulted loans.

The defaulted and non-defaulted loan data can be seen in Figure 5.1. Summary statistics are provided in Table 5.1. There were days when there were 0 defaults with a maximum of 24 defaulted loans in a day. Defaulted loans have a mean value of 7.89 and variance of 25.77. Non-defaulted issued loans ranged from 0 to 31 per day with an average of 8.92 and a variance of 46.29. So, both time series exhibit overdispersion. From the time series plots we see that the data might exhibit a seasonality effect. The sample autocorrelation function, provided in Figure 5.2, also exhibits a seasonal pattern. The partial autocorrelation function (PACF) is provided in Table A.3 of Appendix A. We check the seasonality by marking which days were Saturdays and Sundays in Figure 5.3.

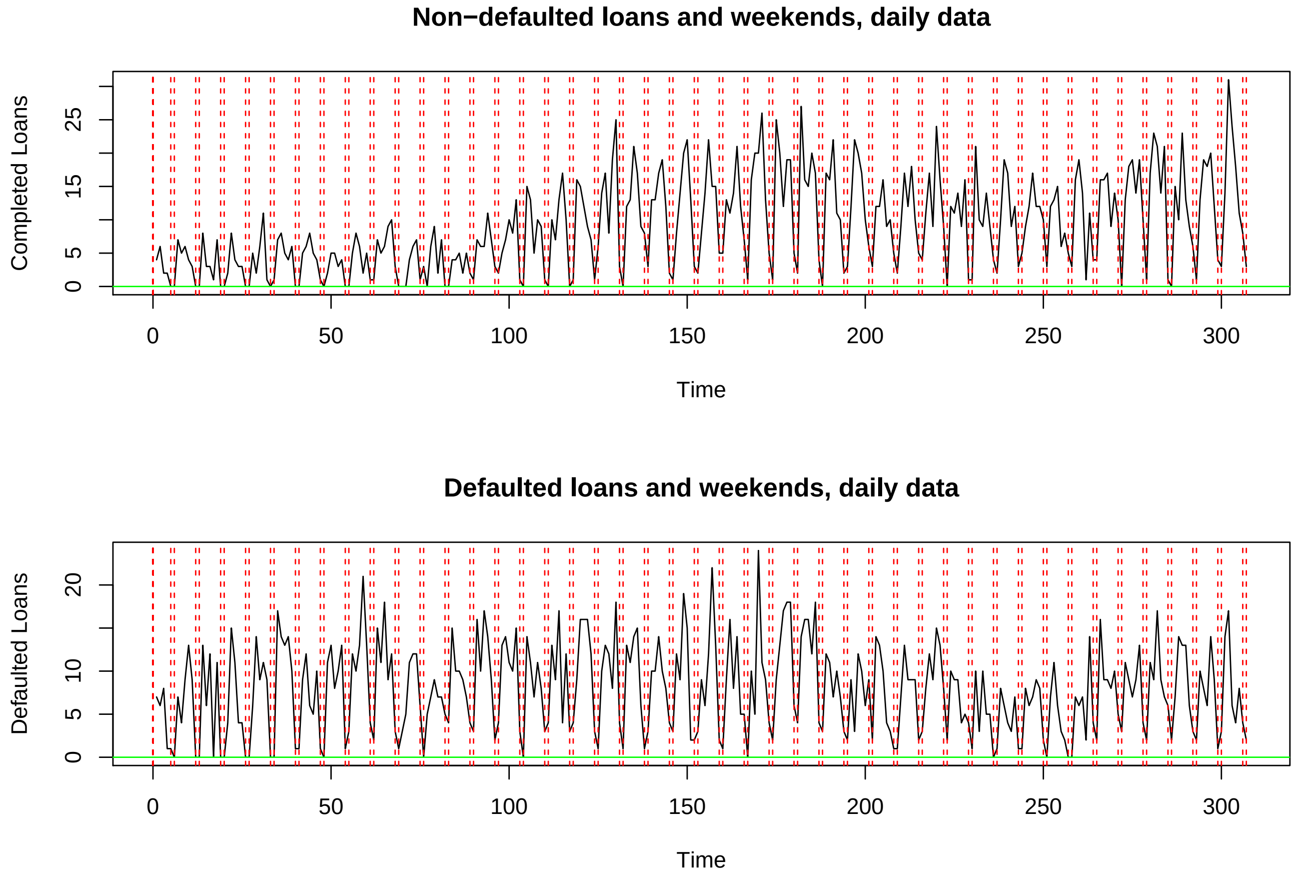


Figure 5.3: Non-defaulted and defaulted loans by their issued date and weekends.

The drops in the amount of issued defaulted loans and non-defaulted loans occurs each Saturday and Sunday, so we conclude that the data exhibits a seasonal effect. After removing the seasonal effect we note that the correlation between the two time series is 0.21089 which indicates that the dependence in the time series could be modelled by a copula.

We want to analyse, whether the amount of loans that have defaulted depends on the amount of loans which have not defaulted but were issued on the same day, so we will consider a BINAR(1) model with different copulas for the innovations. For the marginal distributions of the innovations we will consider Poisson as well as negative binomial distributions. From the Monte Carlo simulation results presented in Section 4.2, we will use the CML estimation method since it has the lowest estimate MSE and bias according to Monte Carlo simulations.

5.2 Estimated models

We estimated a total of 16 BINAR(1) models with different distributions of innovations which include combinations of:

- 4 different copula functions (FGM, Frank, Clayton or Gumbel);
- 4 different combinations of Poisson and negative binomial distributions (both marginals are Poisson, both marginals are negative binomial or a mix of both).

The estimated parameter results are provided in Tables 5.2 through 5.5 with standard errors in parenthesis. $\hat{\alpha}_1, \hat{\lambda}_1, \hat{\sigma}_1^2$ are the parameter estimates of non-defaulted loans and $\hat{\alpha}_2, \hat{\lambda}_2, \hat{\sigma}_2^2$

are parameter estimates of defaulted loans. According to Pawitan (2001), the observed Fisher information is the negative Hessian matrix, evaluated at the MLE. This means, that the standard errors can be numerically derived from the Hessian matrix.

	Copulas with Poisson marginals			
	FGM	Frank	Clayton	Gumbel
$\hat{\alpha}_1$	0.41275	0.41221	0.41029	0.40764
	(0.02441)	(0.02440)	(0.02437)	(0.02471)
$\hat{\alpha}_2$	0.19093	0.19215	0.17919	0.18943
	(0.03846)	(0.03827)	(0.03953)	(0.03889)
$\hat{\lambda}_1$	5.60735	5.61312	5.67114	5.63629
	(0.25376)	(0.25365)	(0.25367)	(0.25604)
$\hat{\lambda}_2$	6.76414	6.75894	6.98387	6.67831
	(0.34644)	(0.34494)	(0.35766)	(0.34847)
$\hat{\theta}$	0.34558	0.94879	0.18532	1.08005
	(0.14577)	(0.35237)	(0.04588)	(0.02875)
Log-likelihood	-1828.7	-1827.712	**-1818.207**	-1823.828
AIC	3667.4	3665.424	**3646.414**	3657.656

Table 5.2: Parameter estimates for BINAR(1) model with Poisson marginals and different copula functions

In Table 5.2 we estimated BINAR(1) models with Poisson marginals and different copula functions. Comparing the log-likelihood and AIC values we see that the Clayton copula provides the best fit for the data. We note that the estimated dependence parameter value for the Clayton copula case is small, indicating weak dependence.

	Copulas with negative binomial marginals			
	FGM	Frank	Clayton	Gumbel
$\hat{\alpha}_1$	0.44064	0.43860	0.43462	0.43574
	(0.03104)	(0.03112)	(0.03101)	(0.03150)
$\hat{\alpha}_2$	0.24248	0.24258	0.23413	0.23163
	(0.04986)	(0.04941)	(0.04978)	(0.04993)
$\hat{\lambda}_1$	5.32757	5.34803	5.39955	5.40970
	(0.37161)	(0.37191)	(0.37123)	(0.37638)
$\hat{\lambda}_2$	6.30057	6.29526	6.39170	6.43102
	(0.44916)	(0.44559)	(0.44880)	(0.44983)
$\hat{\theta}$	0.45976	1.22050	0.26666	1.16689
	(0.18215)	(0.43578)	(0.08731)	(0.05859)
$\hat{\sigma}_1^2$	18.1893	18.27902	18.07136	18.62115
	(2.38261)	(2.39610)	(2.34728)	(2.42026)
$\hat{\sigma}_2^2$	10.844	10.87849	10.80022	11.13189
	(1.16778)	(1.17342)	(1.15866)	(1.19834)
Log-likelihood	-1715.07	-1714.097	**-1710.906**	-1711.95
AIC	3444.14	3442.194	**3435.812**	3437.9

Table 5.3: Parameter estimates for BINAR(1) model with negative binomial marginals and different copula functions

Since the summary statistics of the data sample showed, that the variance of the data is greater than the mean, a negative binomial distribution may provide a better fit. We test this by estimating BINAR(1) models with innovations distributed with negative binomial marginal distributions and different copula functions. The estimation results are provided in Table 5.3.

As in the Poisson marginal distribution case, the Clayton copula provides the best fit for the data. The log likelihood is also smaller compared to the Poisson marginal distribution case. The estimated variance parameters are larger compared to the estimated mean parameters which indicate that overdispersion is indeed present in the data. Other parameter estimates do not differ significantly across different copula functions. However, for the negative binomial distribution case, the estimated mean parameters are larger and the estimated $\hat{\alpha}_j$ parameters are smaller compared to the Poisson marginal distribution case, however, the magnitude of difference is not large, indicating that regardless of selected marginal distributions and copula functions, the estimated values of $\hat{\alpha}_j$ and $\hat{\lambda}_j$ do not exhibit large differences. This also confirms that estimation via CLS is possible for BINAR(1) models, since the estimators of $\hat{\alpha}_j^{CLS}$ and $\hat{\lambda}_j^{CLS}$ do not depend on the distribution of innovations.

	Copulas with Poisson and negative binomial marginals			
	FGM	Frank	Clayton	Gumbel
$\hat{\alpha}_1$	0.41282	0.41246	0.41214	0.40928
	(0.02440)	(0.02438)	(0.02429)	(0.02456)
$\hat{\alpha}_2$	0.24340	0.24305	0.22126	0.23729
	(0.04979)	(0.04951)	(0.05020)	(0.04952)
$\hat{\lambda}_1$	5.59892	5.60049	5.61188	5.65121
	(0.25357)	(0.25323)	(0.25266)	(0.25533)
$\hat{\lambda}_2$	6.33936	6.34615	6.65183	6.27447
	(0.44903)	(0.44689)	(0.45532)	(0.44398)
$\hat{\theta}$	0.38596	0.97657	0.17291	1.06760
	(0.15896)	(0.36825)	(0.05024)	(0.03142)
$\hat{\sigma}_2^2$	10.89328	10.83761	10.52941	10.01072
	(1.17886)	(1.16707)	(1.11914)	(1.07402)
Log-likelihood	-1811.33	-1810.61	**-1805.421**	-1810.831
AIC	3634.66	3633.22	**3622.842**	3633.662

Table 5.4: Parameter estimates for BINAR(1) model with Poisson marginal distribution for non-defaulted loan innovations and negative binomial marginal distribution for defaulted loan innovations and different copula functions

Because copulas can link different marginal distributions it is interesting to see if copulas with different discrete marginal distributions provide a better fit. BINAR(1) model parameter estimates, where non-defaulted loan innovations are modelled with Poisson marginal distributions and defaulted loan innovations are modelled with negative binomial marginal distributions, are provided in Table 5.4. BINAR(1) model parameter estimates where non-defaulted loan innovations are modelled with negative binomial and defaulted loan innovations are modelled with Poisson marginal distributions are provided in Table 5.5. In general, changing one of the marginal distributions to a negative binomial provides a better fit, compared to the Poisson marginal distribution case. As in the negative binomial marginal distribution case, the estimates of α_j and λ_j do not differ significantly compared to the Poisson marginal distribution BINAR(1) models. In both cases, the Clayton copula provides the best fit. Since the variance of non-defaulted loans is higher than defaulted loans, modelling their innovations with a negative binomial marginal distribution provides a better fit than the Poisson case, however, the smallest log-likelihood value is achieved when both marginal distributions are modelled with negative binomial distributions, linked via a Clayton copula.

	Copulas with negative binomial and Poisson marginals			
	FGM	Frank	Clayton	Gumbel
$\hat{\alpha}_1$	0.44003	0.43844	0.43152	0.43541
	(0.03110)	(0.03117)	(0.03128)	(0.03148)
$\hat{\alpha}_2$	0.19015	0.19085	0.18614	0.18295
	(0.03843)	(0.03818)	(0.03856)	(0.03799)
$\hat{\lambda}_1$	5.35566	5.37315	5.50874	5.31839
	(0.37265)	(0.37310)	(0.37562)	(0.36983)
$\hat{\lambda}_2$	6.74058	6.73159	6.78868	6.82293
	(0.34607)	(0.34406)	(0.34712)	(0.34232)
$\hat{\theta}$	0.40179	1.03931	0.21345	1.11175
	(0.16837)	(0.39240)	(0.06612)	(0.04250)
$\hat{\sigma}_1^2$	18.25381	18.28097	17.96076	17.29170
	(2.39244)	(2.39438)	(2.32194)	(2.18606)
Log-likelihood	-1732.608	-1731.83	**-1727.514**	-1730.062
AIC	3477.216	3475.66	**3467.028**	3472.124

Table 5.5: Parameter estimates for BINAR(1) model with negative binomial marginal distribution for non-defaulted loan innovations and Poisson marginal distribution for defaulted loan innovations and different copula functions

Since the Clayton copula provided the best fit, regardless of the selected marginal distributions, it accurately reflects the dependence between non-defaulted and defaulted loan innovations compared to other applied copula functions. From copula descriptions from Section 3, the Clayton copula is used for modelling strong left tail dependence, i.e. when smaller values are more correlated than large values. This would indicate that if there is a small amount of non-defaulted loans, then there is also a small amount of loans that defaulted on the same day. However, the dependence parameter is 0.26666, indicating that the dependence between non-defaulted loans and defaulted loans is weak. We also note that the dependence parameter is relatively close to the sample correlation value which is equal to 0.21089.

6 Conclusions

In this thesis we have analysed different estimation methods for estimating parameters of a BINAR(1) model, including the dependence parameter of its innovations, which are linked via a copula. Monte Carlo simulations were carried out in order to compare different estimation methods. Although estimations of BINAR(1) parameters via CML had the smallest MSE and bias, estimations of the dependence parameter had smaller differences of MSE and bias compared to other estimation methods, indicating that estimations of the dependence parameter via different estimation methods do not exhibit large differences. While CML estimates exhibit the smallest MSE, their estimation via numerical optimization relies on the selection of the initial starting parameter values. These values can be selected via CLS estimation.

An empirical application on loan data was carried out and 16 BINAR(1) models were estimated with different combinations of copula functions and marginal distribution functions. The results showed that, regardless of the selected marginal distributions, out of the 4 copula functions used, the Clayton copula always provided the best model fit. Furthermore, the estimations of α_j and λ_j $(j = 1, 2)$ did not differ significantly throughout the models which indicates that, even a misspecified marginal distribution does not lead to misspecification of these parameters. This also confirms with the CLS estimation method, where α_j and λ_j do not depend on the distribution of the innovations.

Althrough selecting Poisson and negative binomial marginal distribution combinations provided better models compared to models with only Poisson marginal distributions, the models with both marginal distributions modelled via negative binomial distributions provided the smallest log-likelihood values which indicated that both defaulted and non-defaulted loans exhibit overdispersion. The Clayton copula, which provided the best model fit, models variables, which exhibit strong correlation for smaller values and weaker correlation for larger values, however, the estimated dependence parameter is relatively small compared to the size of the interval of possible values it can attain, which indicates that the dependence between defaulted and non defaulted loans is weak. Furthermore, the value of the dependence parameter is similar to the value of the sample correlation between non-defaulted and defaulted loans.

Finally, one can apply different copula functions in order to analyse whether the loan data exhibits different forms of dependence from the ones discussed in this thesis. The asymptotic properties of CLS estimations for the dependence parameter should also be analysed in future research. Lastly, the model can be extended by analysing the presence of structural changes within the data, as well as extending the BINAR(1) model with copula joint innovations to account for the past values of other time series rather than only itself.

References

D. Brigo, A. Pallavicini, and R. Torresetti. *Credit Models and the Crisis: A Journey into CDOs, Copulas, Correlations and Dynamic Models*. Wiley, United Kingdom, 2010.

U. Cherubini, S. Mulinacci, F. Gobbi, and S. Romagnoli. *Dynamic Copula Methods in Finance*. Wiley, United Kingdom, 2011.

J. Crook and F. Moreira. Checking for Asymetric Default Dependence In a Credit Card Portfolio: A Copula Approach. *Journal of Empirical Finance*, 18:728–742, 2011.

J. P. Fenech, H. Vosgha, and S. Shafik. Loan Default Correlation Using an Achimedean Copula Approach: A Case For Recalibration. *Economic Modelling*, 47:340–354, 2015.

C. Genest and J. Nešlehová. A Primer on Copulas for Count Data. *Astin Bulletin*, 37(2): 475–515, 2007.

D. Karlis and X. Pedeli. Flexible Bivariate INAR(1) Processes Using Copulas. *Communications in Statistics - Theory and Methods*, 42:723–740, 2013.

B. Kedem and K. Fokianos. *Regression Models for Time Series Analysis*. Wiley-Interscience, New Jersey, 2002.

R. Nelsen. *An Introduction to Copulas, 2nd edition*. Springer, New York, 2006.

Y. Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, New York, 2001.

X. Pedeli. Modelling Multivariate Time Series for Count Data. *PhD thesis, Athens University of Economis And Business*, 2011.

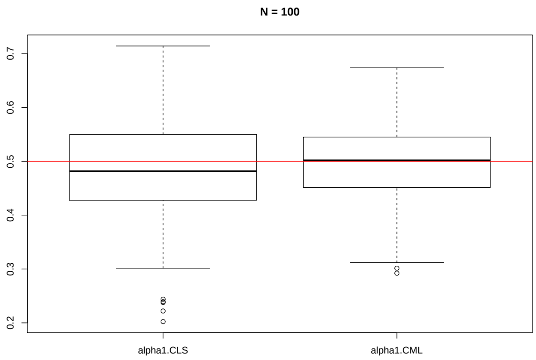
X. Pedeli and D. Karlis. A bivariate INAR(1) process with application. *Statistical Modelling: An International Journal*, 11(4):325–349, 2011.

I. M. M. Silva. Contributions to the analysis of discrete-valued time series. *PhD thesis, University of Porto*, 2005.

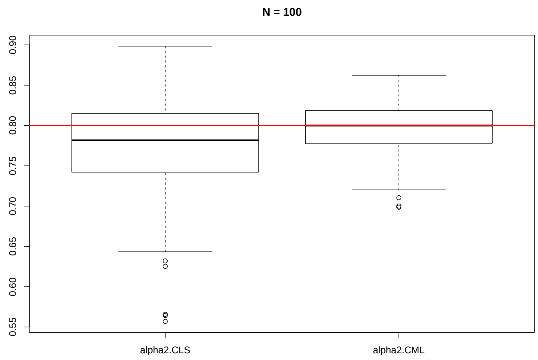
P. K. Trivedi and D. M. Zimmer. Copula Modelling: An Introduction for Practitioners. *Foundations and Trends in Econometrics*, 1(1):1–111, 2007.

A Appendix

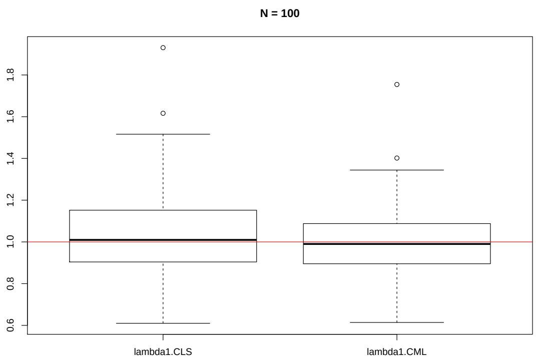
Boxplots of Monte Carlo simulation results for parameters α_j, λ_j, $j = 1, 2$ from Section 4.2.4 with sample size $N = 100$. The interquartile range is smaller for CML estimates and the median is closer to the true parameter value compared to the CLS estimates for all parameters.



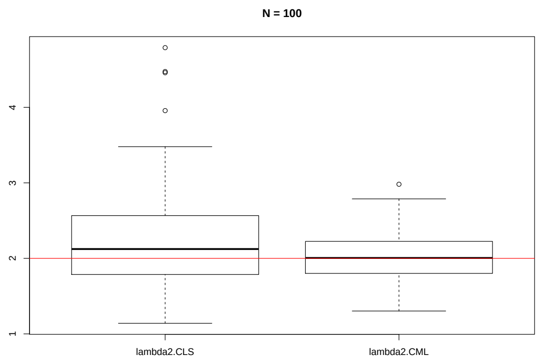
(a) Boxplots of α_1 estimate



(b) Boxplots of α_2 estimate



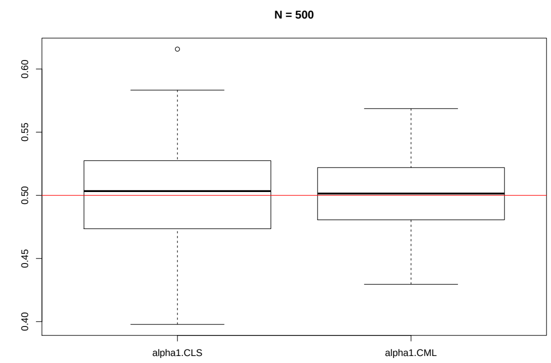
(c) Boxplots of λ_1 estimate



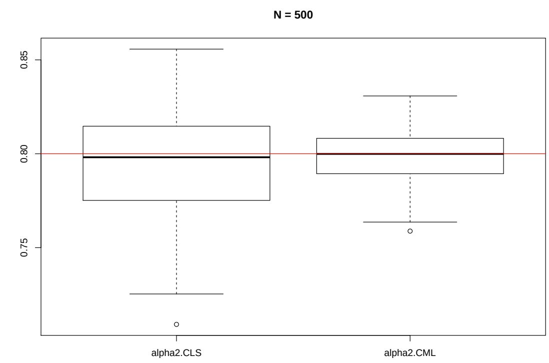
(d) Boxplots of λ_2 estimate

Figure A.1: Monte Carlo simulation results with sample size $N = 100$

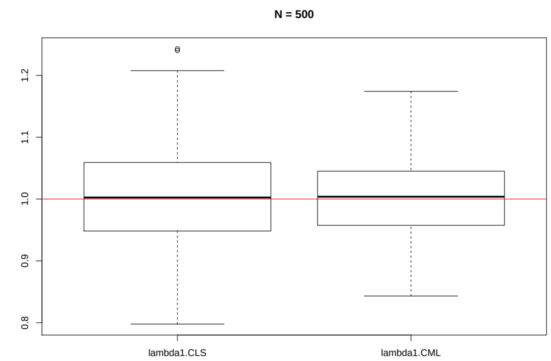
Boxplots of Monte Carlo simulation results for parameters α_j, λ_j, $j = 1, 2$ from Section 4.2.4 with sample size $N = 500$. The median of CLS estimates is closer to the true parameter values compared to the median of CLS estimates for the smaller sample size case. However, the CML estimates are still superior in terms of smaller interquartile range and median estimate values.



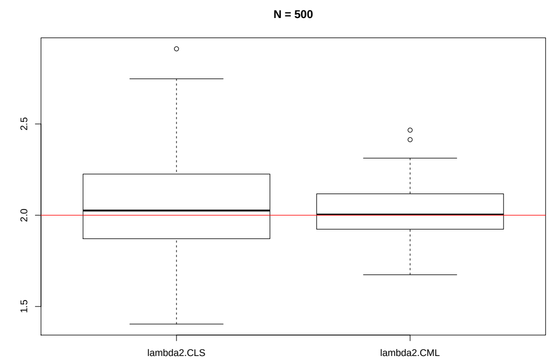
(a) Boxplots of α_1 estimate



(b) Boxplots of α_2 estimate



(c) Boxplots of λ_1 estimate



(d) Boxplots of λ_2 estimate

Figure A.2: Monte Carlo simulation results with sample size $N = 500$

Partial autocorrelation function plot of defaulted and non-defaulted loan data. The first lag is significant, which suggests that the discrete data follows a first-order autoregressive process. Because the data exhibits a seasonality effect (every Saturday and Sunday the amount of issued defaulted and non-defaulted loans is very small compared to other week days), the 6th and 7th lag are also significant.

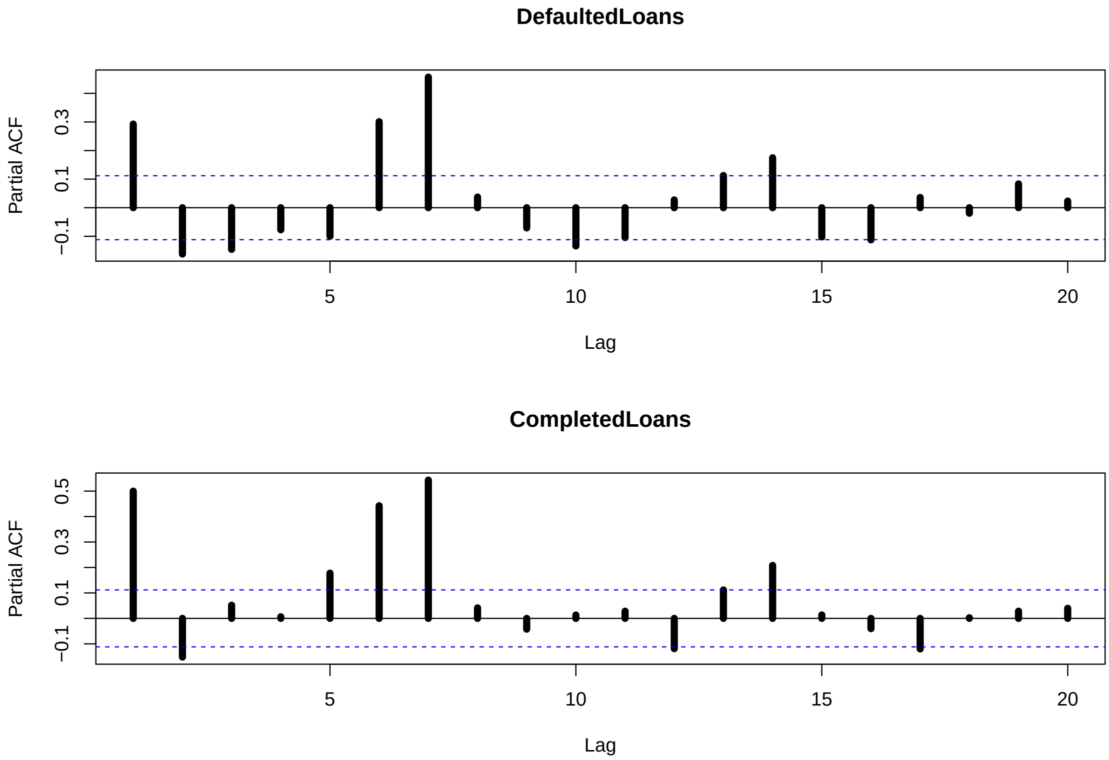


Figure A.3: Partial autocorrelation function plots of defaulted and non-defaulted loan data