https://doi.org/10.15388/vu.thesis.95
https://orcid.org/0000-0003-1723-6004

VILNIUS UNIVERSITY

Andrius
BUTEIKIS

Multivariate copula-based integer-valued time series models: theory and applications

DOCTORAL DISSERTATION

Life sciences,
Mathematics (N 001)

VILNIUS 2020

This dissertation was written between 2016 and 2020 at Vilnius University.

Academic supervisor:
Prof. Habil. Dr. Remigijus Leipus (Vilnius University, Life sciences, Mathematics - N 001).

https://doi.org/10.15388/vu.thesis.95
https://orcid.org/0000-0003-1723-6004

VILNIAUS UNIVERSITETAS

Andrius
BUTEIKIS

Daugiamačiai jungtimis grįstų sveikareikšmių laiko eilučių modeliai: teorija ir taikymai

DAKTARO DISERTACIJA

Gamtos mokslai,
Matematika (N 001)

VILNIUS 2020

Disertacija rengta 2016–2020 metais Vilniaus universitete.

Mokslinis vadovas:
prof. habil. dr. Remigijus Leipus (Vilniaus universitetas, gamtos mokslai, matematika - N 001).

Contents

Preface

Aims and problems

The aim of this dissertation is to analyse autoregressive integer-valued time series models, where the joint innovation distribution is described via a copula. Our objective is the following:

- analyse existing estimation methods for a first-order bivariate integer-valued autoregressive process with copula-joint innovations, and present a two-step estimation method, which allows for a separate estimation of the model parameters from the copula dependence parameter,
- introduce a generalized *univariate* integer-valued autoregressive process for seasonality, where the innovations can be intra-seasonally dependent, which can then be represented in a *multivariate* form that is used for parameter estimation.

Novelty

The results obtained in this thesis extend models in existing literature on integer-valued time series. To the best of our knowledge, as of the writing of this thesis and the resulting publications, the two-step parameter estimation for the model in Chapter 4 and the univariate seasonal specification and its multivariate representation in Chapter 5 were not considered in other papers dealing with integer-valued autoregressive time series. The aforementioned results allowed for faster model parameter estimation, which made estimating such models feasible in empirical applications, which were presented in those chapters. Furthermore, in

Chapter 5 we have introduced a flexible seasonal autoregressive model which allows estimating the seasonal effects in a count data series.

Dissertation structure

This thesis consists of five Chapters, as well as Conclusions, Bibliography and Appendix chapters.

The introduction to integer-valued autoregressive time series models is given in Chapter 1. Existing model specifications are presented.

Chapter 2 presents the main properties of the binomial thinning operator. For the reader's convenience (and because the proofs of some of these properties are not readily available in existing literature), the proofs of these properties are derived and provided in Appendix A.

In Chapter 3 we present the definition of copulas and their applicability to discrete-valued random variables. Four copulas, which are used in this thesis to describe the joint distribution of the model innovations, are provided.

In Chapter 4 we consider bivariate integer-valued autoregressive time series models with copula-joint innovations. Model properties are examined, with proofs provided in Appendix B, and existing parameter estimation methods are considered. We also introduce a two-step parameter estimation method, where the copula parameter is estimated in a separate step. We then compare the accuracy of this method to other existing methods via Monte Carlo simulation. Additional inference on the estimate bias from the simulation is discussed in Appendix C. Futhermore, an empirical application on defaulted and non-defaulted loan counts is carried out.

In Chapter 5 we consider a univariate integer-valued autoregressive process for seasonality with intra-seasonally dependent innovations, where the dependence structure is described via a copula. The specified model allows the autoregressive parameter to vary with the season. Model properties are derived and the proofs are provided in Appendix D. Furthermore, we show that the univariate process can be represented by a multivariate specification, which allows for an efficient parameter estimation in terms of computational speed. Parameter estimation methods are compared via Monte Carlo simulation. Finally, an empirical application is carried out on Chicago crime data to capture the seasonal effects in the series.

Dissemination

The results of this thesis were presented in the following conferences:

- *58th Conference of Lithuanian Mathematical Society*, Vilnius, Lithuania, June 21–22, 2017.
- *20th European Young Statisticians Meeting*, Uppsala, Sweden, August 14–18, 2017.
- *11th International Conference on Computational and Financial Econometrics (CFE 2017) and 10th International Conference of the ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on Computational and Methodological Statistics (CMStatistics 2017)*, London, United Kingdom, December 16–18, 2017.
- *Inaugural Baltic Economic Conference, organized jointly with the 7th Annual Lithuanian Conference on Economic Research*, Vilnius, Lithuania, June 11–12, 2018.
- *12th International Vilnius Conference on Probability Theory and Mathematical Statistics and 2018 IMS Annual Meeting on Probability and Statistics*, Vilnius, Lithuania, July 2–6, 2018.
- *60th Conference of Lithuanian Mathematical Society*, Vilnius, Lithuania, June 19–20, 2019.
- *13th International Conference on Computational and Financial Econometrics (CFE 2019) and 12th International Conference of the ERCIM Working Group on Computational and Methodological Statistics (CMStatistics 2019)*, London, United Kingdom, December 14–16, 2019.

Publications

The results of this thesis are published in the following papers:

- A. Buteikis and R. Leipus. A copula-based bivariate integer-valued autoregressive process with application. *Modern Stochastics: Theory and Applications*, 6(2):227–249, 2019.
- A. Buteikis and R. Leipus. An integer-valued autoregressive process for seasonality. *Journal of Statistical Computation and Simulation*, 90(3):391–411, 2020.

Conference Proceedings

The results of this thesis also appear as abstracts and a short paper in the following conference proceedings:
Short paper:

- A. Buteikis. Copula based BINAR models with applications. *Proceedings of the 20th European young statisticians meeting, Uppsala University, Uppsala, Sweden, 14–18 August 2017*, ISBN: 978-91-506-2680-3.

Abstracts:

- A. Buteikis and R. Leipus. Application of copula-based BINAR models in loan modelling. *11th International Conference on Computational and Financial Econometrics (CFE 2017) and 10th International Conference of the ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on Computational and Methodological Statistics (CMStatistics 2017), London, 16–18 December 2017 : programme and abstracts.*
- A. Buteikis and R. Leipus. Application of copula-based BINAR models in loan modelling. *12th International Vilnius Conference on Probability Theory and Mathematical Statistics and 2018 IMS Annual Meeting on Probability and Statistics, July 2–6, 2018, Vilnius, Lithuania.*

- A. Buteikis and R. Leipus. An integer-valued autoregressive process for seasonality. *13th International Conference on Computational and Financial Econometrics (CFE 2019) and 12th International Conference of the ERCIM Working Group on Computational and Methodological Statistics (CMStatistics 2019), London, 14–16 December 2019: programme and abstracts.*

Acknowledgements

I would like to express my sincere gratitude to my PhD supervisor, prof. Remigijus Leipus for his patience and advice in answering the variety of questions, which arose during the course of my PhD studies and research. I would also like to thank assoc. prof. Martynas Manstavičius and prof. Márton Ispány, for their attentive remarks and invaluable suggestions, as well as assoc. prof. Viktor Skorniakov and dr. Mátyás Barczy for their insights and observations for improving the dissertation.

And to my parents and brother, for their support and encouragement.

Chapter 1

Introduction

Integer-valued time series comprising count observations at regular time intervals can be observed in various applications, such as the number of stock trades per day, the amount of crimes committed in a city per hour, the number of worker strikes in a country per month, the amount of insurance claims in a firm per year, the number of defaulted loans issued by a bank per week, the number of infected people per day, etc.

Usually a company's methods for evaluating loan risk are not publicly available, however, one way to measure whether insolvent clients are adequately separated from responsible clients would be to look at the quantity of defaulted and non-defaulted issued loans each day. The adequacy of a firms rules for issuing loans can be analysed by modelling the dependence between the number of loans which have defaulted and number of loans that have not defaulted via copulas.

The advantage of such approach is that copulas allow to model the marginal distributions (possibly 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 survival analysis, hydrology, insurance risk analysis as well as finance (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.

The dependence between the default rate of loans between different credit risk categories were analysed in Crook and Moreira (2011). In order to model the dependence, copulas from ten 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 global 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) discussed the differences and challenges of using copulas for discrete data compared to continuous data. Note that the previously mentioned studies assumed that the data does not depend on its own previous values.

One of the simplest models for individual evolution is the integer-valued autoregressive model of order one (INAR(1)), independently proposed by Al-Osh and Alzaid (1987) and McKenzie (1986), which is defined as

$$Y_t = \phi \circ Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{1.0.1}$$

where $\phi \circ Y_{t-1} = \sum_{i=1}^{Y_{t-1}} B_{i,t}$, where, for all $t \in \mathbb{Z}$, $\{B_{i,t}, i \in \mathbb{Z}\}$ is a sequence of i.i.d. Bernoulli random variables (r.v.s) with mean $\phi \in [0,1]$, such that these sequences are mutually independent and independent of the sequence of i.i.d. nonnegative integer-valued r.v.s $\{\varepsilon_t, t \in \mathbb{Z}\}$. ε_t are independent of Y_{t-k} for $k > 0$. The popular choices for distribution of ε_t are Poisson or negative-binomial.

By using bivariate integer-valued autoregressive models (BINAR) it is possible to account for both the discreteness and autocorrelation of the data when analysing a pair of time series count 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.

On the other hand, many such time series observations display a seasonal phenomenon, which may arise from various periodic (daily, weekly, yearly, etc.) factors. The apparent seasonal patterns can be described by using, e.g., explanatory variables. Brännäs (1995) introduced explanatory variables to the INAR(1) process by allowing the parameters in (1.0.1) to be time-varying. The INAR(1) process with explanatory process can be written as

$$Y_t = \phi_t \circ Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim \text{Pois}(\lambda_t), \tag{1.0.2}$$

where ϕ_t represents the survival probability and, in economic context, can be thought of as the survival probability of a firm and thus depends on the business cycle phase. λ_t represents the mean entry of a random shock in the time series, which could be interpreted as the mean number of firms entering the market at each period t. Brännäs (1995) assumed that ϕ_t is the logistic function (thus $0 < \phi_t < 1$) and λ_t is the exponential function:

$$\phi_t = \frac{1}{1+\exp(\mathbf{X}_t^\top \beta)}, \quad \lambda_t = \exp(\mathbf{Z}_t^\top \gamma),$$

where $\mathbf{X}_t$, $\mathbf{Z}_t$ are fixed explanatory variable vectors and β, γ are unknown parameter vectors. This specification allows to specify the seasonality and trend as explanatory variables, which are introduced through the autoregressive coefficient and the mean coefficient of the random component distribution. Ding and Wang (2016) proposed the maximum empirical likelihood estimator for the parameters of the INAR(1) process with explanatory variables and demonstrated, via a simulation study, the latter to be slightly better than the conditional least squares estimation method introduced by Brännäs (1995). The model in eq. (1.0.2) was later expanded to a higher order INAR process by Enciso-Mora et al. (2009).

An INAR process with a periodic structure was proposed by Monteiro et al. (2010). They defined a periodic INAR(1) process with period d (PINAR$(1)_d$) as

$$Y_t = \phi_t \circ Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{N}, \quad \varepsilon_t \sim \text{Pois}(\nu_t), \tag{1.0.3}$$

where $\phi_t = \alpha_j \in (0,1)$ and $\nu_t = \lambda_j$ for $t = j + kd$, $j = 1, \ldots, d$, $k \in \mathbb{N}_0 := \{0,1,2,\ldots\}$ and $\phi_t \circ Y_{t-1} = \sum_{i=1}^{Y_{t-1}} B_{i,t}$, where $\{B_{i,t} = B_{i,t}(\phi_t), t \in \mathbb{Z}\}$ is a periodic sequence of Bernoulli random variables independent of Y_t and ε_t, with $\mathbb{P}(B_{i,t}(\phi_t) = 1) = \phi_t$. Furthermore, in Monteiro et al. (2010) the ε_t are assumed to be a periodic sequence of independent Poisson random variables, independent of Y_{t-1} and $\phi_t \circ Y_{t-1}$. In eq. (1.0.3) the periodic structure was introduced through both the autoregressive and the shock components of the model. Because of this structure, a dimensionality reduction in the number of parameters is mentioned as one of the possible challenges.

A seasonal structure for an INAR(1) Poisson process was imposed by Bourguignon et al. (2016), who introduced a first-order seasonal INAR process with seasonality d (INAR$(1)_d$) of the following form

$$Y_t = \phi \circ Y_{t-d} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{1.0.4}$$

where $\phi \in [0,1]$. ε_t and $\phi \circ Y_{t-d}$ are defined in the same way as in (1.0.1). The INAR$(1)_d$ process consists of mutually independent INAR(1) processes.

Furthermore, there exist various extensions, which replace the binomial thinning operator in eq. (1.0.1) with a more general thinning operator. Such thinning operators are presented in Section 2.3. Alternatively, Barreto-Souza (2017) proposed a Mixed Poisson INAR(1) process for modelling overdispersed count time series. Overdispersion is commonly observed in count time series and can be caused by an excess of zeros or unobserved significant covariates. The proposed process is defined via the binomial thinning operator if $Y_t \sim \mathrm{MP}(\mu, \phi)$, $\forall t \in \mathbb{Z}$ and:

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{1.0.5}$$

where $\alpha \in [0,1)$ and $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of i.i.d. r.v.s, independent of Y_s, for $s < t$, $\forall t \in \mathbb{Z}$. Unlike the process defined in eq. (1.0.1), Y_t is now assumed to follow a mixed Poisson distribution. In general, a random variable Y follows a mixed Poisson distribution, if it satisfies the stochastic representation $Y|W = w \sim Pois(\mu w)$ for $\mu > 0$, where W is some non-negative random variable.

Barreto-Souza (2017) indicated that one of the challenges of analysing the process in eq. (1.0.5) is that it is very difficult to obtain an explicit form for the probability function of the innovations, ε_t, and in their specific case, the innovations satisfy the stochastic representation of $\varepsilon|W = w \sim Pois(\mu w)$, where W has the distribution of an innovation of an inverse-Gaussian AR(1) process, which does not have an explicit density function.

Consequently, we focus on integer-valued autoregressive processes, defined with the binomial thinning operator.

Chapter 2

Properties of the binomial thinning operator

We present the main properties of the binomial thinning operator, which will be used when defining BINAR(1) and $\text{SINAR}(1)_d$ processes in Chapter 4 and Chapter 5. Denote by '$\overset{d}{=}$' the equality of distributions.

The aforementioned binomial thinning operator was introduced by Steutel and van Harn (1979). A survey of various thinning operators, which generalize the binomial thinning operator is presented in Weiß (2008). In this thesis, we focus on models with the binomial thinning operator and leave the analysis of other thinning operators as future extensions for these models.

2.1 Binomial thinning operator properties for the univariate case

A number of binomial thinning operator properties are provided in Pedeli (2011) and Silva (2005). Along with the main thinning operator properties, we also define some additional properties in the following theorem.

Theorem 2.1.1. ***Binomial thinning operator properties****. Let X, X_1, X_2 be non-negative integer-valued random variables, such that $\mathbb{E}(Z^2) < \infty$, $Z \in \{X, X_1, X_2\}$, $\alpha, \alpha_1, \alpha_2 \in [0,1)$ and let '$\circ$' be the binomial thinning operator, such that $\beta \circ Z = \sum_{i=1}^{Z} B_i$ for any $\beta \in \{\alpha, \alpha_1, \alpha_2\}$ and Z, where B_i are independent Bernoulli random variables, independent of*

Z, with $\mathbb{P}(B_i = 1) = \beta$. Then the following properties hold:

(a) $\alpha_1 \circ (\alpha_2 \circ X) \overset{d}{=} (\alpha_1 \alpha_2) \circ X$*;*

(b) $\alpha \circ (X_1 + X_2) \overset{d}{=} \alpha \circ X_1 + \alpha \circ X_2$*;*

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

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

(e) $\mathbb{E}(X_2^p(\alpha \circ X_1)) = \alpha \mathbb{E}(X_1 X_2^p)$*, with* $p \geq 0$*, assuming* $\mathbb{E}(X_1 X_2^p) < \infty$*;*

(f) $\mathbb{E}(X^p(\alpha \circ X)) = \alpha \mathbb{E}(X^{p+1})$*, with* $p \geq 0$*, assuming* $\mathbb{E}(X^{p+1}) < \infty$*;*

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

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

(i) $\mathbb{E}(X^p(\alpha \circ X)^2) = \alpha^2 \mathbb{E}(X^{p+2}) + \alpha(1-\alpha)\mathbb{E}(X^{p+1})$*, with* $p \geq 0$*, assuming* $\mathbb{E}(X^{p+2}) < \infty$*.*

Complete proofs are derived and provided in Appendix A.1.

2.2 Binomial thinning operator properties for the multivariate case

The univariate binomial thinning operator can be generalized to the multivariate setting with the following theorem.

Theorem 2.2.1. ***Binomial thinning operator properties in the multivariate setting.*** *Let* $\mathbf{X}_j = [X_{1,j}, \ldots, X_{k,j}]^\top$*,* $j = 1, 2$ *be non-negative integer-valued random k-vectors and let* $\mathbf{A}_j = (\alpha_{i_1,i_2,j})_{i_1,i_2=1,\ldots,k}$*, with* $\alpha_{i_1,i_2,j} \in [0,1)$*,* $\forall i_1, i_2 = 1, \ldots, k$ *and* $j = 1, 2$*. In the multivariate setting the binomial thinning operator also acts as a matrix multiplication and is defined by*

$$\mathbf{A}_j \circ \mathbf{X}_j = \left[\sum_{s=1}^k \alpha_{1,s,j} \circ X_{s,j} \quad , \ \ldots \ , \quad \sum_{s=1}^k \alpha_{k,s,j} \circ X_{s,j} \right]^\top$$

with the following properties:

(a) $\mathbb{E}[\mathbf{A}_j \circ \mathbf{X}_j] = \mathbf{A}_j \mathbb{E}[\mathbf{X}_j]$*;*

(b) $\mathbb{E}[\mathbf{A}_i \circ \mathbf{X}_i][\mathbf{A}_j \circ \mathbf{X}_j]^\top = \mathbf{A}_i \mathbb{E}[\mathbf{X}_i \mathbf{X}_j^\top]\mathbf{A}_j^\top + \mathbb{1}_{\{i=j\}} diag(\mathbf{B}_i \mathbb{E}[\mathbf{X}_i])$, *where* $\mathbf{B}_i = (\alpha_{m_1,m_2,i}(1-\alpha_{m_1,m_2,i}))_{m_1,m_2=1,\dots,k}$ *is the covariance matrix of the independent Bernoulli random variables from* $\alpha_{m,m,i} \circ X_{m,i} = \sum_{l=1}^{X_{m,i}} B_{m,m,i,l}$, $\forall m = 1,\dots,k$, *independent of* $X_{m,i}$, *with* $\mathbb{E}(B_{m,m,i,l}) = \alpha_{m,m,i}$, $\mathbb{V}\text{ar}(B_{m,m,i,l}) = \alpha_{m,m,i}(1-\alpha_{m,m,i})$, $\forall m \in \{1,\dots,k\}$, *and* $i,j = 1,2$;

(c) $\mathbb{E}[\mathbf{A}_i \circ \mathbf{X}_i][\mathbf{A}_j \mathbf{X}_j]^\top = \mathbf{A}_i \mathbb{E}[\mathbf{X}_i \mathbf{X}_j^\top]\mathbf{A}_j^\top$;

(d) $\mathbb{E}[\mathbf{A}_i \mathbf{X}_i][\mathbf{A}_j \circ \mathbf{X}_j]^\top = \mathbf{A}_i \mathbb{E}[\mathbf{X}_i \mathbf{X}_j^\top]\mathbf{A}_j^\top$.

The proofs of properties (a) and (b) are provided in Latour (1997). Complete proofs are derived in Appendix A.2.

2.3 Other thinning operators

The focus of this thesis is on models based on the binomial thinning operator, however, alternative thinning operators are available. In this section we will briefly present a select few alternative thinning operators to highlight the various different ways that the binomial thinning operator is generalized in literature.

One possible extension of the binomial thinning operator is the generalized thinning operator, which was proposed by Latour (1998) and is defined for a non-negative integer-valued r.v. X as:

$$\alpha \bullet_\beta X := \sum_{j=1}^{X} Z_j,\ \alpha \in [0,1],$$

where $Z_j \in \mathbb{N} \cup \{0\}$ are i.i.d. r.v.s and independent of X, with $\mathbb{E}(Z_j) = \alpha$ and $\mathbb{V}\text{ar}(Z_j) = \beta$. The properties of the generalized thinning operator are similar to the binomial thinning operator: $\mathbb{E}(\alpha \bullet_\beta X) = \alpha \mathbb{E}(X)$, $\mathbb{V}\text{ar}(\alpha \bullet_\beta X) = \alpha^2 \mathbb{V}\text{ar}(X) + \beta \mathbb{E}(X)$, $\mathbb{C}\text{ov}(\alpha \bullet_\beta X, X) = \alpha \mathbb{V}\text{ar}(X)$.

Another modification of the binomial thinning operator, introduced by Kim and Park (2006), generalizes to allow for both non-negative and negative integers:

$$\alpha \odot X := \text{sgn}(\alpha)\text{sgn}(X) \sum_{j=1}^{|X|} B_j,\ \alpha \in (-1,1),$$

where $X \in \mathbb{Z}$. $\{B_j, j \in \mathbb{Z}\}$ is a sequence of i.i.d. Bernoulli r.v.s with $\mathbb{P}(B_j = 1) = |\alpha|$, $\forall j = 1,...,X$, $\text{sgn}(X) = 1$, if $X \geq 0$ and $\text{sgn}(X) = -1$ otherwise. If $X \in \mathbb{N} \cup \{0\}$ and $\alpha \geq 0$, then the signed binomial thinning is reduced to the binomial thinning.

The negative binomial thinning operator was introduced by Ristić et al. (2009) and is defined as:

$$\alpha * X := \sum_{j=1}^{X} W_j, \ \alpha \in [0,1),$$

where $\{W_j, j \in \mathbb{Z}\}$ is a sequence of i.i.d. $\text{Geometric}(\alpha/(1+\alpha))$ r.v.s independent of X. The negative binomial thinning operator describes a geometric counting series, which can help explain overdispersion.

In order to model count time series with equidispersion, underdiserpsion and overdispersion Bourguignon and Weiß (2017) defined the BiNB thinning operator for a non-negative integer-valued r.v. X as:

$$(\alpha,\beta) \circledast X := \sum_{j=1}^{X} W_j, \ \alpha+\beta \in [0,1),$$

where $\{W_j, j \in \mathbb{Z}\}$ is a sequence of i.i.d. $\text{BerG}(\alpha,\beta)$, independent of X. The BerG distribution is a convolution of the Bernoulli and geometric distribution, i.e. the sum $W := Z_1 + Z_2$, where $Z_1 \sim Bern(\alpha)$ and $Z_2 \sim \text{Geometric}(1/(1+\beta))$ are independent with $\mathbb{E}(Z_1) = \alpha \in (0,1)$ and $\mathbb{E}(Z_2) = \beta > 0$, has the following probability mass function is defined as:

$$\mathbb{P}(W = w) = \begin{cases} \dfrac{1-\alpha}{1+\beta} & , \text{ if } w = 0, \\ (\alpha+\beta)\dfrac{\beta^{w-1}}{(1+\beta)^{w+1}} & , \text{ if } w \in \mathbb{N}, \end{cases}$$

so that $\mathbb{E}(W) = \alpha + \beta$ and $\mathbb{V}\text{ar}(Z) = \alpha(1-\alpha) + \beta(1+\beta)$.

Lastly, a new thinning operator based on the Gómez–Déniza–Sarabia–Calderín-Ojeda (GSC) distribution was proposed by Kang et al. (2020). The probability mass function of the $\text{GSC}(\alpha, \ \theta)$ distribution is defined as:

$$\mathbb{P}(W = w) = \frac{\log(1-\alpha\theta^{w}) - \log(1-\alpha\theta^{w+1})}{\log(1-\alpha)},$$

where $\alpha < 1$, $\alpha \neq 0$ and $0 < \theta < 1$. The GSC thinning operator is defined as:

$$\alpha \diamond X := \sum_{j=1}^{X} W_j, \ \alpha < 1, \ \alpha \neq 0,$$

where $\{W_j, j \in \mathbb{Z}\}$ is a sequence of i.i.d. GSC$(\alpha, \ \exp\{-|\alpha|\})$ r.v.s, independent of X. This thinning operator can capture equidispersion, overdispersion and underdispersion of the series, as well as zero-inflated, zero-deflated, short-tailed and long-tailed characteristics of the count data.

For a detailed survey of various thinning operators for discrete data time series models, see Weiß (2008), Scotto et al. (2015), Davis et al. (2016), Weiß (2018), Joe (2019) and the references therein.

Chapter 3

Count random variable dependence and copulas

In this section we review 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.

Copulas are used for modelling the dependence between several random variables. The main advantage of using copulas is that they allow to model the marginal distributions separately from their joint distribution. A two-dimensional copula, which is used in Chapter 4, is defined below.

Definition 3.0.1. A 2-dimensional copula $C : [0,1]^2 \to [0,1]$ is a function with the following properties:

(i) for every $u, v \in [0,1]$:

$$C(u,0) = C(0,v) = 0; \tag{3.0.1}$$

(ii) for every $u, v \in [0,1]$:

$$C(u,1) = u, \quad C(1,v) = v; \tag{3.0.2}$$

(iii) for any $u_1, u_2, v_1, v_2 \in [0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$:

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0 \tag{3.0.3}$$

(this is also called *the rectangle inequality*, *2-increasingness property*, or *supermodularity* of bivariate functions).

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

Theorem 3.0.1. [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 for all $(x_1, x_2) \in [-\infty, \infty]^2$:*

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \tag{3.0.4}$$

If F_i is continuous for $i = 1, 2$ then C is unique; otherwise C is uniquely determined only on $\text{Range}(F_1) \times \text{Range}(F_2)$*, where* $\text{Range}(F)$ *denotes the range of the cdf F. Conversely, if C is a copula and F_1, F_2 are distribution functions, then the function H, defined by equation* (3.0.4) *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. More about Copula theory, properties and applications can be found in Nelsen (2006) and Joe (2015).

Copulas with discrete marginal distributions

Since the innovations of models in Chapter 4 and Chapter 5 are nonnegative integer-valued random variables, one needs to consider copulas for constructing multivariate distributions with discrete marginals F_1 and F_2 in eq. (3.0.4). In this section we will mention some of the key differences when copula marginals are discrete rather than continuous.

Firstly, as mentioned in Theorem 3.0.1, if F_1 and F_2 are discrete marginals then a unique copula representation exists only for values in the range of $\text{Range}(F_1) \times \text{Range}(F_2)$. However, the lack of uniqueness does not pose a problem in empirical applications because it implies that there may exist more than one copula which describes the distribution of the empirical data. Secondly, regarding concordance and discordance, the discrete case has to allow for ties (i.e. when two variables have the same value), so the concordance measures (Spearman's rho and Kendal's tau) are margin-dependent, see Trivedi and Zimmer (2007). There are several modifications proposed for Spearman's rho, however, none of them are margin-free. Furthermore, Genest and Nešlehová (2007) state that estimators of a copula family parameter based on Kendall's tau or its modified versions are biased and estimation techniques based on maximum likelihood are recommended. As such, we will not examine estimation methods based on concordance measures. Another difference from the continuous case is the use of the probability mass function (pmf) instead of the probability density function when estimating the model parameters which will be seen in Section 4.2.

Finally, we note that in some cases biased estimators may be more desirable if one can control the bias and has a computationally fast procedure for the biased estimator (e.g. Kendal's tau). For a discussion on biased versus unbiased estimation, see Efron (1975). In addition to the bias problem, Genest and Nešlehová (2007, Section 6.2) state that rank-based methods do not always lead to a consistent estimator of the copula parameter when the marginals are discrete.

Some concrete copulas

In this section we will present several bivariate copulas, which will be used later when constructing and evaluating the BINAR(1) model in Chapter 4. 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 following Genest and Nešlehová (2007), the parameter of a copula family, denoted θ, is referred to as the dependence parameter in the context of this thesis, if a corresponding copula family is increasing in concordance order.

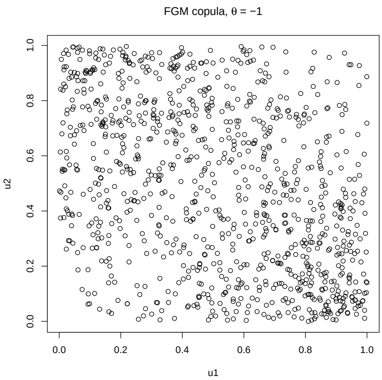
Farlie-Gumbel-Morgenstern copula

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

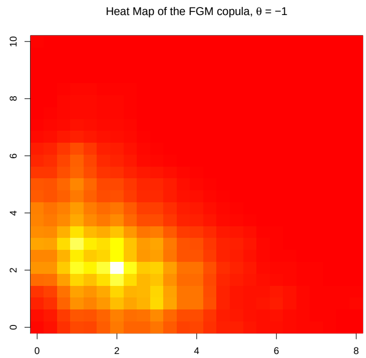
$$C(u_1,u_2;\theta) = u_1u_2(1+\theta(1-u_1)(1-u_2)). \tag{3.0.5}$$

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)).

Figure 3.0.1 shows the FGM copula, when the dependence parameter is -1 and the marginal cdfs F_1, F_2 are either of continuous or discrete random variables. The heat map is used to visualize the relationship between two integer-valued series (for example, see Weiß (2018, Figure 2.4)). We can see that even when θ has the minimum value, the dependence between the random variables isn't strong.



(a) Continuous marginals



(b) Discrete marginals

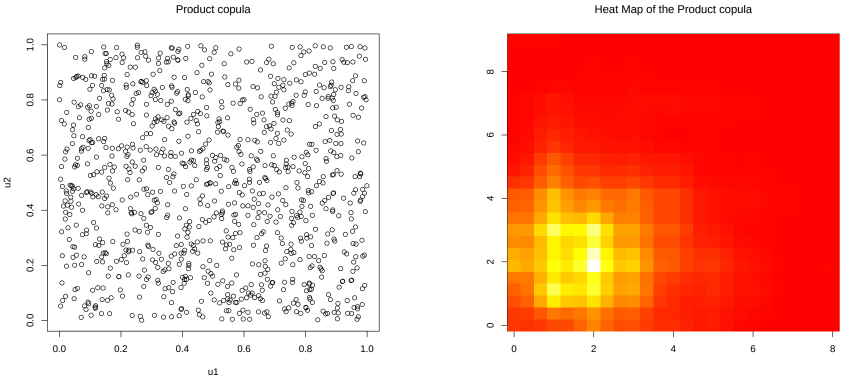
Figure 3.0.1: The FGM copula for the continuous marginal case and the discrete marginal case.

Product copula

The copula when $\theta = 0$ is called a product (or independence) copula:

$$C(u_1, u_2) = u_1 u_2. \tag{3.0.6}$$

Since the product copula corresponds to independence, it is important as a benchmark.



(a) Continuous marginals

(b) Discrete marginals

Figure 3.0.2: The product copula for the continuous marginal case and the discrete marginal case.

Figure 3.0.2 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.

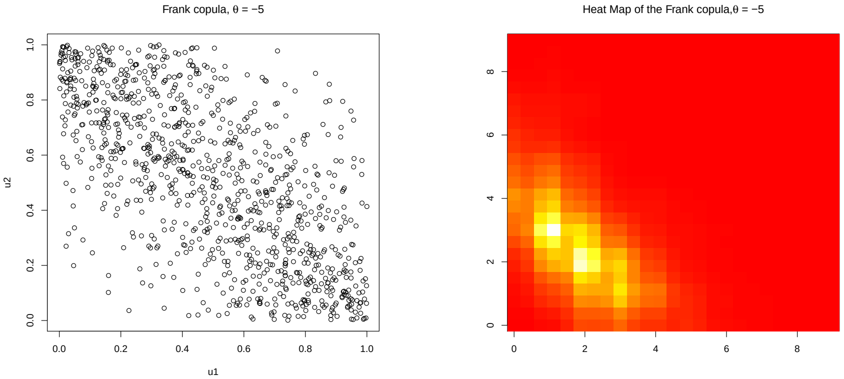
Frank copula

The Frank copula has the following form:

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

The dependence parameter can take values from $(-\infty,\infty)\setminus\{0\}$. The Frank copula allows for both positive and negative dependence[1] between the marginals.

[1]Let $\mathbf{X}=(X_1,X_2)$ be a bivariate random vector with cdf F. Then F, or $\mathbf{X}$, is positive (quadrant) dependent if $\mathbb{P}(X_1 > a_1, X_2 > a_2) \geq \mathbb{P}(X_1 > a_1)\mathbb{P}(X_2 > a_2), \forall a_1, a_2 \in \mathbb{R}$. If the inequality is reversed, then F, or $\mathbf{X}$, is negative (quadrant) dependent. Similar conditions apply to the general multivariate case. See Joe (2015, Section 2.8).



(a) Continuous marginals **(b)** Discrete marginals

Figure 3.0.3: The Frank copula for the continuous marginal case and the discrete marginal case.

Figure 3.0.3 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.

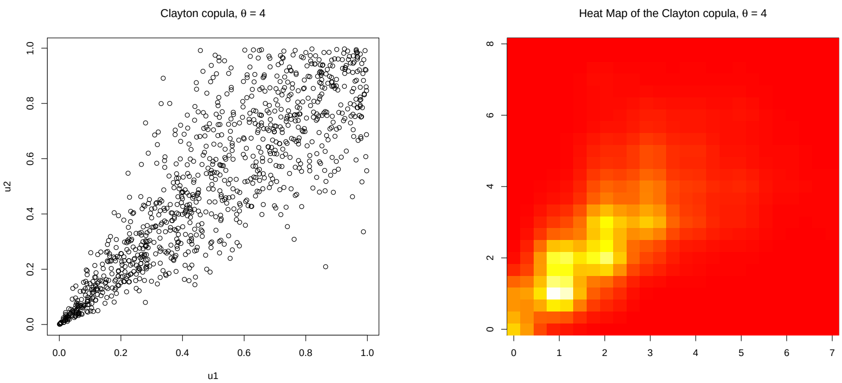
Clayton copula

The Clayton copula has the following form:

$$C(u_1, u_2; \theta) = \max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\}^{-\frac{1}{\theta}}, \tag{3.0.7}$$

with the dependence parameter $\theta \in [-1, \infty) \setminus \{0\}$. The marginals become independent when $\theta \to 0$. It can be used when the correlation between two random variables exhibits a strong left tail dependence – if smaller values are strongly correlated and higher values are less correlated [2]. The Clayton copula can also account for negative dependence when $\theta \in [-1, 0)$. For more properties of this copula, see recent paper Manstavičius and Leipus (2017).

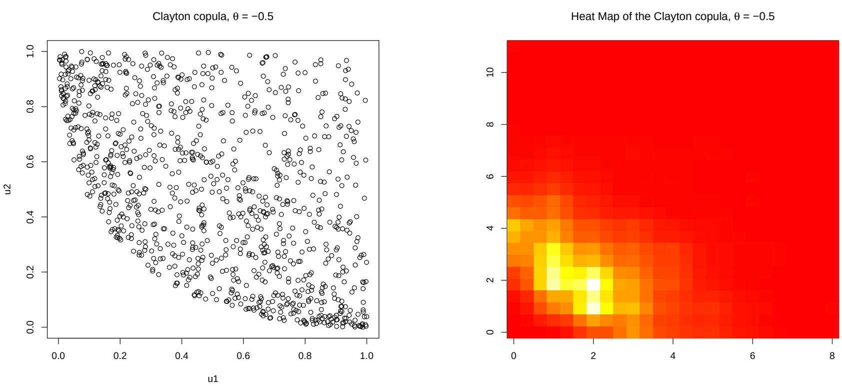
[2]If a bivariate copula C is such that $\lim_{u \to 1^-} \overline{C}(u,u)/(1-u) = \lambda_U$ exists, then C has upper tail dependence if $\lambda_U \in (0, 1]$. Similarly, if $\lim_{u \to 0^+} C(u,u)/u = \lambda_L$ exists, then C has lower tail dependence if $\lambda_L \in (0, 1]$. Here $\overline{C}$ is the survival function of the copula. See Joe (2015, Section 2.13).



(a) Continuous marginals **(b)** Discrete marginals

Figure 3.0.4: The Clayton copula for the continuous marginal case and the discrete marginal case.

Figure 3.0.4 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.



(a) Continuous marginals **(b)** Discrete marginals

Figure 3.0.5: The Clayton copula for the continuous marginal case and the discrete marginal case with a negative dependence parameter.

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 3.0.5.

Furthermore, the above copulas can be extended to a higher dimension case. The joint distribution of $[u_1, \ldots, u_d]^\top$, denoted $F(\cdot)$, is described by a d-dimensional copula C - a distribution function $C : [0,1]^d \to [0,1]$ with uniform margins, such that

$$F(a_1, ..., a_d) = C(F_1(a_1), ..., F_d(a_d)),$$

where $F_j(\cdot)$ is the univariate marginal distribution of u_j, $j \in \{1, ..., d\}$. For discrete random variables, the copula is uniquely determined only on $\mathrm{Range}(F_1) \times ... \times \mathrm{Range}(F_d)$. Below we present the multivariate versions for copulas, which are used in Chapter 5.

d-variate Frank copula

The d-variate version of the Frank copula is given by

$$C(u_1, ..., u_d; \theta) = -\frac{1}{\theta} \log \left(1 + \frac{\prod_{i=1}^{d} [\exp(-\theta u_i) - 1]}{[\exp(-\theta) - 1]^{d-1}} \right),$$

with $\theta > 0$.

d-variate Clayton copula

The d-variate version of the Clayton copula is given by

$$C(u_1, ..., u_d; \theta) = \left[\sum_{i=1}^{d} u_i^{-\theta} - d + 1 \right]^{-\frac{1}{\theta}},$$

with $\theta > 0$.
A compendium of various different copulas for both the bivariate and d-variate cases can be found in Nadarajah et al. (2018).

Chapter 4

A copula-based bivariate integer-valued autoregressive process

In this chapter we expand on using copulas in BINAR models by analysing additional copula families for the innovations of the BINAR(1) model and analyse different BINAR(1) model parameter estimation methods. We also present a two-step 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 (including the one used in Karlis and Pedeli (2013)) are compared 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 for empirical weekly loan data.

4.1 The bivariate INAR(1) process

The BINAR(1) process was introduced in Pedeli and Karlis (2011). In this section we will provide the definition of the BINAR(1) model and will formulate its properties.

Definition 4.1.1. Let $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \varepsilon_{2,t}]^\top$, $t \in \mathbb{Z}$, be a sequence of independent identically distributed (i.i.d.) nonnegative integer-valued bivariate random variables. A bivariate integer-valued autoregressive process of order 1 (BINAR(1)), $\mathbf{Y}_t = [Y_{1,t}, Y_{2,t}]^\top$, $t \in \mathbb{Z}$, is defined as:

$$\mathbf{Y}_t = \mathbf{A} \circ \mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad t \in \mathbb{Z}, \tag{4.1.1}$$

where $\alpha_j \in [0,1)$, $j = 1,2$, and the symbol '$\circ$' is the thinning operator which also acts as the matrix multiplication. So the jth ($j = 1,2$) element is defined as an INAR process of order 1 (INAR(1)):

$$Y_{j,t} = \alpha_j \circ Y_{j,t-1} + \varepsilon_{j,t}, \quad t \in \mathbb{Z}, \tag{4.1.2}$$

where $\alpha_j \circ Y_{j,t-1} := \sum_{i=1}^{Y_{j,t-1}} B_{j,t,i}$ and $B_{j,t,1}, B_{j,t,2}, \ldots$ is a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(B_{j,t,i} = 1) = \alpha_j = 1 - \mathbb{P}(B_{j,t,i} = 0)$, $\alpha_j \in [0,1)$, such that these sequences are mutually independent and independent of the sequence $\boldsymbol{\varepsilon}_t$, $t \in \mathbb{Z}$.

Consequently, for each t, $\boldsymbol{\varepsilon}_t$ is independent of $\mathbf{Y}_s$, $s < t$. $Y_{j,t}$, defined in eq. (4.1.2), 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 $\alpha_j \circ Y_{j,t-1}$, and the elements which enter in the system in the interval $(t-1, t]$, which are called arrival elements and denoted by $\varepsilon_{j,t}$. We can obtain an infinite series representation by substitutions and the properties of the thinning operator as in Al-Osh and Alzaid (1987) or Kedem and Fokianos (2002, p. 180):

$$Y_{j,t} = \alpha_j \circ Y_{j,t-1} + \varepsilon_{j,t} \stackrel{d}{=} \sum_{k=0}^{\infty} \alpha_j^k \circ \varepsilon_{j,t-k}, \quad j = 1,2, \quad t \in \mathbb{Z}, \tag{4.1.3}$$

where convergence on the right-hand side holds a.s. The properties of the binomial thinning operator are provided in Chapter 2.

Now we present some properties of the BINAR(1) model. They will be used when analysing some of the parameter estimation methods.

Theorem 4.1.1. [Properties of the BINAR(1)) process] *Let* $\mathbf{Y}_t = [Y_{1,t}, Y_{2,t}]^\top$ *be a nonnegative integer-valued time series given in Def. 4.1.1 and* $\alpha_j \in [0,1)$, $j = 1,2$. *Let* $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \varepsilon_{2,t}]^\top$, $t \in \mathbb{Z}$, *be nonnegative integer-valued random variables with* $\mathbb{E}(\varepsilon_{j,t}) = \mu_{\varepsilon,j}$ *and* $\mathbb{V}\mathrm{ar}(\varepsilon_{j,t}) = \sigma_j^2$, $j = 1,2$. *Then the following properties hold:*

(a) $\mathbb{E}Y_{j,t} = \mu_{Y_j} = \frac{\mu_{\varepsilon,j}}{1-\alpha_j}$;

(b) $\mathbb{E}(Y_{j,t}|Y_{j,t-1}) = \alpha_j Y_{j,t-1} + \mu_{\varepsilon,j}$;

(c) $\mathbb{V}\text{ar}(Y_{j,t}) = \sigma_{Y_j}^2 = \dfrac{\sigma_j^2 + \alpha_j \mu_{\varepsilon,j}}{1-\alpha_j^2}$;

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

(e) $\mathbb{C}\text{ov}(Y_{j,t}, Y_{j,t+h}) = \alpha_j^h \sigma_{Y_j}^2,\ h \geq 0$;

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

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

(h) $\mathbb{C}\text{orr}(Y_{i,t+h}, Y_{j,t}) = \dfrac{\alpha_i^h \sqrt{(1-\alpha_i^2)(1-\alpha_j^2)}}{(1-\alpha_i\alpha_j)\sqrt{(\sigma_i^2+\alpha_i\mu_{\varepsilon,i})(\sigma_j^2+\alpha_j\mu_{\varepsilon,j})}} \mathbb{C}\text{ov}(\varepsilon_{i,t}, \varepsilon_{j,t})$,

$i \neq j,\ h \geq 0$;

The proofs for these properties can be easily derived and a number of these are provided in Pedeli (2011). For the reader's convenience, the proofs of the above properties are also provided in Appendix B.1.

Similarly to (4.1.3), it holds that

$$\mathbf{Y}_t \stackrel{d}{=} \sum_{k=0}^{\infty} \mathbf{A}^k \circ \boldsymbol{\varepsilon}_{t-k},$$

where convergence on the right-hand side holds a.s.

Hence, the distributional properties of the BINAR(1) process can be studied in terms of $\boldsymbol{\varepsilon}_t$ values. Note also, that according to Latour (1997), if $\alpha_j \in [0,1),\ j = 1,2$, then there exists a unique stationary nonnegative integer-valued sequence $\mathbf{Y}_t$, $t \in \mathbb{Z}$, satisfying (4.1.1).

From the covariance (g) and correlation (h) of the BINAR(1) process we see that the dependence between $Y_{1,t}$ and $Y_{2,t}$ depends on the joint distribution of the innovations $\varepsilon_{1,t},\ \varepsilon_{2,t}$. 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 distributions of innovations of a BINAR(1) model are linked by either a Frank copula or a normal copula with either Poisson or negative binomial marginal distributions. 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.

4.2 Parameter estimation of the copula-based BINAR(1) model

In this section we examine different BINAR(1) model parameter estimation methods and provide a two-step estimation method for estimating the copula dependence parameter separately from the other parameters. Estimation methods are compared via Monte Carlo simulations. Let $\mathbf{Y}_t = [Y_{1,t}, Y_{2,t}]^\top$ be a non-negative integer-valued time series given in Def. 4.1.1, where the joint distribution of $[\varepsilon_{1,t}, \varepsilon_{2,t}]^\top$, with marginals F_1, F_2, is linked by a copula $C(\cdot,\cdot)$:

$$\mathbb{P}(\varepsilon_{1,t} \leq y_1, \varepsilon_{2,t} \leq y_2) = C(F_1(y_1), F_2(y_2))$$

and let $C(u_1,u_2) = C(u_1,u_2;\theta)$, where θ is a dependence parameter.

4.2.1 Conditional least squares estimation

The Conditional least squares (CLS) estimator minimizes the squared distance between $\mathbf{Y}_t$ and its conditional expectation. Similarly to the method in Silva (2005) for the INAR(1) model, we construct the CLS estimator in the case of the BINAR(1) model.
Using Theorem 2.1.1 we can write the vector of conditional means as

$$\boldsymbol{\mu}_{t|t-1} := \begin{bmatrix} \mathbb{E}(Y_{1,t}|Y_{1,t-1}) \\ \mathbb{E}(Y_{2,t}|Y_{2,t-1}) \end{bmatrix} = \begin{bmatrix} \alpha_1 Y_{1,t-1} + \mu_{\varepsilon,1} \\ \alpha_2 Y_{2,t-1} + \mu_{\varepsilon,2} \end{bmatrix}, \tag{4.2.1}$$

where $\mu_{\varepsilon,j} := \mathbb{E}\varepsilon_{j,t},\ j = 1,2$. In order to calculate the CLS estimators of $(\alpha_1, \alpha_2, \mu_{\varepsilon,1}, \mu_{\varepsilon,2})$, we define the vector of residuals as the difference between the observations and their conditional expectation:

$$\mathbf{Y}_t - \boldsymbol{\mu}_{t|t-1} = \begin{bmatrix} Y_{1,t} - \alpha_1 Y_{1,t-1} - \mu_{\varepsilon,1} \\ Y_{2,t} - \alpha_2 Y_{2,t-1} - \mu_{\varepsilon,2} \end{bmatrix}.$$

Then, given a sample of N observations, $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$, the CLS estimators of $\alpha_j, \mu_{\varepsilon,j},\ j = 1,2$ are found by minimizing the sum

$$Q_j(\alpha_j, \mu_{\varepsilon,j}) := \sum_{t=2}^{N} (Y_{j,t} - \alpha_j Y_{j,t-1} - \mu_{\varepsilon,j})^2 \longrightarrow \min_{\alpha_j, \mu_{\varepsilon,j}}, \quad j = 1,2.$$

By taking the derivatives with respect to α_j and $\mu_{\varepsilon,j},\ j = 1,2$ and equating them to zero we get:

$$\widehat{\alpha}_j^{\mathrm{CLS}} = \frac{\sum_{t=2}^{N} (Y_{j,t} - \bar{Y}_j)(Y_{j,t-1} - \bar{Y}_j^*)}{\sum_{t=2}^{N} (Y_{j,t-1} - \bar{Y}_j^*)^2} \tag{4.2.2}$$

and

$$\widehat{\mu}_{\varepsilon,j}^{\mathrm{CLS}} = \frac{1}{N-1} \left(\sum_{t=2}^{N} Y_{j,t} - \widehat{\alpha}_j^{\mathrm{CLS}} \sum_{t=2}^{N} Y_{j,t-1} \right), \tag{4.2.3}$$

where $\bar{Y}_j := (N-1)^{-1} \sum_{t=2}^{N} Y_{j,t}$ and $\bar{Y}_j^* := (N-1)^{-1} \sum_{t=2}^{N} Y_{j,t-1}$ (see Ispány et al. (2003)). The asymptotic properties of the CLS estimators for the INAR(1) model case are provided in Latour (1998), Silva (2005), Barczy et al. (2010) and can be applied to the BINAR(1) parameter estimates, specified via equations (4.2.2) and (4.2.3). By the fact that the j-th component of the BINAR(1) process is an INAR(1) itself, we can formulate the following theorem for the marginal parameter vector distributions (see Barczy et al. (2010)):

Theorem 4.2.1. *Let $\mathbf{Y}_t = [Y_{1,t}, Y_{2,t}]^\top$ be defined in Def. 4.1.1 and let the parameter vector of (4.1.2) be $[\alpha_j, \mu_{\varepsilon,j}]^\top$. Assume that $\widehat{\alpha}_j^{\mathrm{CLS}}$ and $\widehat{\mu}_{\varepsilon,j}^{\mathrm{CLS}}$ are the* CLS *estimators of α_j and $\mu_{\varepsilon,j},\ j = 1,2$. Then:*

$$\sqrt{N} \begin{pmatrix} \widehat{\alpha}_j^{\mathrm{CLS}} - \alpha_j \\ \widehat{\mu}_{\varepsilon,j}^{\mathrm{CLS}} - \mu_{\varepsilon,j} \end{pmatrix} \xrightarrow{d} \mathscr{N}(\mathbf{0}_2, \mathbf{B}_j),$$

where

$$\mathbf{B}_j = \begin{bmatrix} \mathbb{E}Y_{j,t}^2 & \mathbb{E}Y_{j,t} \\ \mathbb{E}Y_{j,t} & 1 \end{bmatrix}^{-1} \mathbf{A}_j \begin{bmatrix} \mathbb{E}Y_{j,t}^2 & \mathbb{E}Y_{j,t} \\ \mathbb{E}Y_{j,t} & 1 \end{bmatrix}^{-1},$$

$$\mathbf{A}_j = \alpha_j(1-\alpha_j) \begin{bmatrix} \mathbb{E}Y_{j,t}^3 & \mathbb{E}Y_{j,t}^2 \\ \mathbb{E}Y_{j,t}^2 & \mathbb{E}Y_{j,t} \end{bmatrix} + \sigma_j^2 \begin{bmatrix} \mathbb{E}Y_{j,t}^2 & \mathbb{E}Y_{j,t} \\ \mathbb{E}Y_{j,t} & 1 \end{bmatrix}, \ j = 1,2.$$

Here, the moments are derived in Barczy et al. (2010) as

$$
\begin{aligned}
\mathbb{E}Y_{j,t} &= \frac{\mu_{\varepsilon,j}}{1-\alpha_j}, \quad \mathbb{E}Y_{j,t}^2 = \frac{\sigma_j^2+\alpha_j\mu_{\varepsilon,j}}{1-\alpha_j^2} + \frac{\mu_{\varepsilon,j}^2}{(1-\alpha_j)^2},\\
\mathbb{E}Y_{j,t}^3 &= \frac{\mathbb{E}\varepsilon_{j,t}^3 - 3\sigma_j^2(1+\mu_{\varepsilon,j}) - \mu_{\varepsilon,j}^3 + 2\mu_{\varepsilon,j}}{1-\alpha_j^3} + 3\frac{\sigma_j^2+\alpha_j\mu_{\varepsilon,j}}{1-\alpha_j^2} - 2\frac{\mu_{\varepsilon,j}}{1-\alpha_j}\\
&\quad + 3\frac{\mu_{\varepsilon,j}(\sigma_j^2+\alpha_j\mu_{\varepsilon,j})}{(1-\alpha_j)(1-\alpha_j^2)} + \frac{\mu_{\varepsilon,j}^3}{(1-\alpha_j)^3}.
\end{aligned}
$$

For the Poisson marginal distribution case the asymptotic variance matrix can be expressed as (see Freeland and McCabe (2005))

$$
\mathbf{B}_j = \begin{bmatrix} \frac{\alpha_j(1-\alpha_j)^2}{\mu_{\varepsilon,j}} + 1 - \alpha_j^2 & -(1+\alpha_j)\mu_{\varepsilon,j} \\ -(1+\alpha_j)\mu_{\varepsilon,j} & \mu_{\varepsilon,j} + \frac{1+\alpha_j}{1-\alpha_j}\mu_{\varepsilon,j}^2 \end{bmatrix}, \; j = 1,2.
$$

Furthermore, for a more general case, Latour (1997) proved that the CLS estimators of a multivariate generalized integer-valued autoregressive process (GINAR) are asymptotically normally distributed.

Note that

$$
\mathbb{E}(Y_{1,t} - \alpha_1 Y_{1,t-1} - \mu_{\varepsilon,1})(Y_{2,t} - \alpha_2 Y_{2,t-1} - \mu_{\varepsilon,2}) = \mathbb{C}\mathrm{ov}(\varepsilon_{1,t}, \varepsilon_{2,t}), \qquad (4.2.4)
$$

which follows from

$$
\begin{aligned}
&\mathbb{E}(Y_{1,t} - \alpha_1 Y_{1,t-1} - \mu_{\varepsilon,1})(Y_{2,t} - \alpha_2 Y_{2,t-1} - \mu_{\varepsilon,2})\\
&\quad = \mathbb{E}(\alpha_1 \circ Y_{1,t-1} - \alpha_1 Y_{1,t-1})(\alpha_2 \circ Y_{2,t-1} - \alpha_2 Y_{2,t-1})\\
&\quad + \mathbb{E}(\alpha_1 \circ Y_{1,t-1} - \alpha_1 Y_{1,t-1})(\varepsilon_{2,t} - \mu_{\varepsilon,2})\\
&\quad + \mathbb{E}(\alpha_2 \circ Y_{2,t-1} - \alpha_2 Y_{2,t-1})(\varepsilon_{1,t} - \mu_{\varepsilon,1})\\
&\quad + \mathbb{E}(\varepsilon_{1,t} - \mu_{\varepsilon,1})(\varepsilon_{2,t} - \mu_{\varepsilon,2})
\end{aligned}
$$

since the first three summands are zeros.

Example 4.2.1. Assume that joint probability mass function of $[\varepsilon_{1,t}, \varepsilon_{2,t}]$ is given by a bivariate Poisson distribution:

$$
\mathbb{P}(\varepsilon_{1,t} = k, \varepsilon_{2,t} = l) = \sum_{i=0}^{\min\{k,l\}} \frac{(\mu_{\varepsilon,1} - \lambda)^{k-i}(\mu_{\varepsilon,2} - \lambda)^{l-i}\lambda^i}{(k-i)!(l-i)!i!} \mathrm{e}^{-(\mu_{\varepsilon,1} + \mu_{\varepsilon,2} - \lambda)},
$$

where $\mu_{\varepsilon,j} > 0$, $j = 1,2$, $0 \le \lambda < \min\{\mu_{\varepsilon,1}, \mu_{\varepsilon,2}\}$. Then, for each $j = 1,2$, the marginal distribution of $\varepsilon_{j,t}$ is Poisson with parameter $\mu_{\varepsilon,j}$ and $\mathbb{C}\mathrm{ov}(\varepsilon_{1,t}, \varepsilon_{2,t}) = \lambda$. If $\lambda = 0$ then the two variables are independent.

Example 4.2.2. Assume that joint probability mass function of $[\varepsilon_{1,t}, \varepsilon_{2,t}]$ is a bivariate negative binomial distribution given by

$$\mathbb{P}(\varepsilon_{1,t} = k, \varepsilon_{2,t} = l) = \frac{\Gamma(\beta+k+l)}{\Gamma(\beta)k!l!}\Big(\frac{\mu_{\varepsilon,1}}{\mu_{\varepsilon,1}+\mu_{\varepsilon,2}+\beta}\Big)^k\Big(\frac{\mu_{\varepsilon,2}}{\mu_{\varepsilon,1}+\mu_{\varepsilon,2}+\beta}\Big)^l \times \Big(\frac{\beta}{\mu_{\varepsilon,1}+\mu_{\varepsilon,2}+\beta}\Big)^\beta,$$

where $\mu_{\varepsilon,j} > 0$, $j = 1,2$, $\beta > 0$. Then, for each $j = 1,2$, the marginal distribution of $\varepsilon_{j,t}$ is negative binomial with parameters β and $p_j = \beta/(\mu_{\varepsilon,j}+\beta)$ and $\mathbb{E}\varepsilon_{j,t} = \mu_{\varepsilon,j}$, $\mathbb{V}\mathrm{ar}(\varepsilon_{j,t}) = \mu_{\varepsilon,j}(1+\beta^{-1}\mu_{\varepsilon,j})$, $\mathbb{C}\mathrm{ov}(\varepsilon_{1,t}, \varepsilon_{2,t}) = \beta^{-1}\mu_{\varepsilon,1}\mu_{\varepsilon,2}$. Thus, bivariate negative binomial distribution is more flexible than bivariate Poisson due to overdispersion parameter β.

Assume now that the Poisson innovations $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ with parameters $\mu_{\varepsilon,1}$ and $\mu_{\varepsilon,2}$, respectively, are joint by a copula with dependence parameter θ. Taking into account equality (4.2.4), we can estimate θ by minimizing the sum of squared differences

$$S = \sum_{t=2}^{N}\big(\varepsilon_{1,t}^{\mathrm{CLS}}\varepsilon_{2,t}^{\mathrm{CLS}} - \gamma(\hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}}, \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}};\theta)\big)^2, \tag{4.2.5}$$

where

$$\varepsilon_{j,t}^{\mathrm{CLS}} := Y_{j,t} - \hat{\alpha}_j^{\mathrm{CLS}}Y_{j,t-1} - \hat{\mu}_{\varepsilon,j}^{\mathrm{CLS}}, \quad j = 1,2$$

and

$$\begin{aligned}\gamma(\mu_{\varepsilon,1}, \mu_{\varepsilon,2};\theta) &:= \mathbb{C}\mathrm{ov}(\varepsilon_{1,t}, \varepsilon_{2,t}) \\ &= \sum_{k,l=1}^{\infty} kl\, c(F_1(k;\mu_{\varepsilon,1}), F_2(l;\mu_{\varepsilon,2});\theta) - \mu_{\varepsilon,1}\mu_{\varepsilon,2}.\end{aligned} \tag{4.2.6}$$

Here, $c(F_1(k;\mu_{\varepsilon,1}),F_2(l;\mu_{\varepsilon,2});\theta)$ is the joint probability mass function:

$$\begin{aligned} c(F_1(k;\mu_{\varepsilon,1}),F_2(l;\mu_{\varepsilon,2});\theta) &= \mathbb{P}(\varepsilon_{1,t}=k,\varepsilon_{2,t}=l) \\ &= C(F_1(k;\mu_{\varepsilon,1}),F_2(l;\mu_{\varepsilon,2});\theta) \\ &- C(F_1(k-1;\mu_{\varepsilon,1}),F_2(l;\mu_{\varepsilon,2});\theta) \\ &- \ C(F_1(k;\mu_{\varepsilon,1}),F_2(l-1;\mu_{\varepsilon,2});\theta) \\ &+ \ C(F_1(k-1;\mu_{\varepsilon,1}),F_2(l-1;\mu_{\varepsilon,2});\theta), \quad (4.2.7) \\ & k\geq 1,\ l\geq 1. \end{aligned}$$

Our estimation method is based on the approximation of covariance $\gamma(\hat{\mu}_{\varepsilon,1}^{\text{CLS}},\hat{\mu}_{\varepsilon,2}^{\text{CLS}};\theta)$ by

$$\gamma^{(M_1,M_2)}(\hat{\mu}_{\varepsilon,1}^{\text{CLS}},\hat{\mu}_{\varepsilon,2}^{\text{CLS}};\theta) = \sum_{k=1}^{M_1}\sum_{l=1}^{M_2} kl\, c(F_1(k;\hat{\mu}_{\varepsilon,1}^{\text{CLS}}),F_2(l;\hat{\mu}_{\varepsilon,2}^{\text{CLS}});\theta) - \hat{\mu}_{\varepsilon,1}^{\text{CLS}}\hat{\mu}_{\varepsilon,2}^{\text{CLS}}. \tag{4.2.8}$$

For example, if the marginals are Poisson with parameters $\mu_{\varepsilon,1}=\mu_{\varepsilon,2}=1$ and their joint distribution is given by FGM copula in (3.0.5), then the covariance $\gamma^{(M_1,M_2)}(1,1;\theta)$ stops changing significantly after setting $M_1=M_2=M=8$, regardless of the selected dependence parameter θ. We used this approximation methodology when carrying out a Monte Carlo simulation in Section 4.3.

For the FGM copula, if we take the derivative of the sum

$$S^{(M_1,M_2)} = \sum_{t=2}^{N}\left(\varepsilon_{1,t}^{\text{CLS}}\varepsilon_{2,t}^{\text{CLS}} - \gamma^{(M_1,M_2)}(\hat{\mu}_{\varepsilon,1}^{\text{CLS}},\hat{\mu}_{\varepsilon,2}^{\text{CLS}};\theta)\right)^2, \tag{4.2.9}$$

equate it to zero and use equation (4.2.8), we get

$$\hat{\theta}^{\text{FGM}} = \frac{\sum_{t=2}^{N}(Y_{1,t}-\hat{\alpha}_1^{\text{CLS}}Y_{1,t-1}-\hat{\mu}_{\varepsilon,1}^{\text{CLS}})(Y_{2,t}-\hat{\alpha}_2^{\text{CLS}}Y_{2,t-1}-\hat{\mu}_{\varepsilon,2}^{\text{CLS}})}{(N-1)\sum_{k=1}^{M_1}k(F_{1,k}\overline{F}_{1,k}-F_{1,k-1}\overline{F}_{1,k-1})\sum_{l=1}^{M_2}l(F_{2,l}\overline{F}_{2,l}-F_{2,l-1}\overline{F}_{2,l-1})}, \tag{4.2.10}$$

where $F_{j,k} := F_j(k;\hat{\mu}_{\varepsilon,j}^{\text{CLS}})$, $\overline{F}_{j,k} := 1-F_{j,k}$, $j=1,2$. The derivation of equation (4.2.10) is straightforward and is presented in Appendix B.2 for the reader's convenience.

Depending on the selected copula family, calculating (4.2.7) to get the analytical expression of estimator $\hat{\theta}$ may be difficult. However, we can use the function `optim` in R to minimize (4.2.5). For other marginal distribution cases, where the marginal distribution has parameters other than expected value $\mu_{\varepsilon,j}$, equation (4.2.5) would need to be minimized by those additional parameters. For example, in the case of negative binomial marginals with corresponding mean $\mu_{\varepsilon,j}$ and variance σ_j^2, i.e. if

$$\mathbb{P}(\varepsilon_{j,t}=k)=\frac{\Gamma\Big(k+\frac{\mu_{\varepsilon,j}^2}{\sigma_j^2-\mu_{\varepsilon,j}}\Big)}{\Gamma\Big(\frac{\mu_{\varepsilon,j}^2}{\sigma_j^2-\mu_{\varepsilon,j}}\Big)k!}\left(\frac{\mu_{\varepsilon,j}}{\sigma_j^2}\right)^{\frac{\mu_{\varepsilon,j}^2}{\sigma_j^2-\mu_{\varepsilon,j}}}\left(\frac{\sigma_j^2-\mu_{\varepsilon,j}}{\sigma_j^2}\right)^k,$$

$k=0,1,\ldots,\ j=1,2$, then the additional parameters are σ_1^2,σ_2^2 and minimization problem becomes

$$S^{(M_1,M_2)}\longrightarrow\min_{\sigma_1^2,\sigma_2^2,\theta}.$$

4.2.2 Conditional maximum likelihood estimation

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

$$\begin{aligned}
&\mathbb{P}(Y_{1,t}=y_{1,t},Y_{2,t}=y_{2,t}|Y_{1,t-1}=y_{1,t-1},Y_{2,t-1}=y_{2,t-1})\\
&\quad=\mathbb{P}(\alpha_1\circ y_{1,t-1}+\varepsilon_{1,t}=y_{1,t},\alpha_2\circ y_{2,t-1}+\varepsilon_{2,t}=y_{2,t})\\
&\quad=\sum_{k=0}^{y_{1,t}}\sum_{l=0}^{y_{2,t}}\mathbb{P}(\alpha_1\circ y_{1,t-1}=k)\mathbb{P}(\alpha_2\circ y_{2,t-1}=l)\mathbb{P}(\varepsilon_{1,t}=y_{1,t}-k,\varepsilon_{2,t}=y_{2,t}-l).
\end{aligned}$$

Here, $\alpha_j\circ y$ is the sum of y independent Bernoulli trials. Hence,

$$\mathbb{P}(\alpha_j\circ y_{j,t-1}=k)=\binom{y_{j,t-1}}{k}\alpha_j^k(1-\alpha_j)^{y_{j,t-1}-k},\quad k=0,\ldots,y_{j,t-1},\ j=1,2.$$

In the case of copula-based BINAR(1) model with Poisson marginals,

$$\mathbb{P}(\varepsilon_{1,t}=y_{1,t}-k,\varepsilon_{2,t}=y_{2,t}-l)=c(F_1(y_{1,t}-k,\mu_{\varepsilon,1}),F_2(y_{2,t}-l,\mu_{\varepsilon,2});\theta).$$

Thus, we obtain

$$
\begin{aligned}
&\mathbb{P}(Y_{1,t}=y_{1,t},Y_{2,t}=y_{2,t}|Y_{1,t-1}=y_{1,t-1},Y_{2,t-1}=y_{2,t-1})\\
&=\sum_{k=0}^{y_{1,t}}\sum_{l=0}^{y_{2,t}}\binom{y_{1,t-1}}{k}\alpha_1^k(1-\alpha_1)^{y_{1,t-1}-k}\binom{y_{2,t-1}}{l}\alpha_2^l(1-\alpha_2)^{y_{2,t-1}-l}\\
&\quad\times c(F_1(y_{1,t}-k,\mu_{\varepsilon,1}),F_2(y_{2,t}-l,\mu_{\varepsilon,2});\theta)
\end{aligned}
$$

and the log conditional likelihood function, for estimating the marginal distribution parameters $\mu_{\varepsilon,1},\mu_{\varepsilon,2}$, the probabilities of the Bernoulli trial successes α_1,α_2 and the dependence parameter θ, is

$$
\begin{aligned}
&\ell(\alpha_1,\alpha_2,\mu_{\varepsilon,1},\mu_{\varepsilon,2},\theta)\\
&=\sum_{t=2}^{N}\log\mathbb{P}(Y_{1,t}=y_{1,t},Y_{2,t}=y_{2,t}|Y_{1,t-1}=y_{1,t-1},Y_{2,t-1}=y_{2,t-1})
\end{aligned}
$$

for some initial values $y_{1,1}$ and $y_{2,1}$. In order to estimate the unknown parameters we maximize the log conditional likelihood:

$$
\ell(\alpha_1,\alpha_2,\mu_{\varepsilon,1},\mu_{\varepsilon,2},\theta)\longrightarrow\max_{\alpha_1,\alpha_2,\mu_{\varepsilon,1},\mu_{\varepsilon,2},\theta}. \tag{4.2.11}
$$

Numerical maximization is straightforward with the `optim` function in R statistical software.

As with the CLS estimator case, for other marginal distribution cases where the marginal distribution has parameters other than $\mu_{\varepsilon,j}$, equation (4.2.11) would need to be maximized by those additional parameters. The CML estimator is asymptotically normally distributed under standard regularity conditions and its variance matrix is the inverse of the Fisher information matrix (see Pedeli and Karlis (2011)).

4.2.3 Two-step estimation based on CLS and CML

Depending on the range of attainable values of the parameters and the sample size, CML maximization might take some time to compute. On the other hand, since CLS estimators of α_j and $\mu_{\varepsilon,j}$ 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. (4.2.11) with CLS estimates from equations (4.2.2) and (4.2.3). Then we will only need to maximize ℓ with respect to a single dependence parameter θ for the Poisson marginal distribution case.

In other words, the two-step approach for estimating the unknown parameters consists of finding

$$(\hat{\alpha}_j^{\mathrm{CLS}}, \hat{\mu}_{\varepsilon,j}^{\mathrm{CLS}}) = \arg\min Q_j(\alpha_j, \mu_{\varepsilon,j}), \quad j = 1,2$$

in the first step and taking those values as given in the second step:

$$\hat{\theta}^{\mathrm{CML}} = \arg\max \ell(\hat{\alpha}_1^{\mathrm{CLS}}, \hat{\alpha}_2^{\mathrm{CLS}}, \hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}}, \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}}, \theta).$$

For other marginal distribution cases, any additional parameters, other than α_j and $\mu_{\varepsilon,j}$ would be estimated in the second step.

4.3 Estimation method comparison via Monte Carlo simulation

We carried out a Monte Carlo simulation 1000 times to test the estimation methods with sample size 50 and 500. The generated model was a BINAR(1) with innovations joint by either an FGM, Frank or Clayton copula with Poisson marginal distributions as well as the case when the marginal distributions can be from different families: one is a Poisson and the other – a negative binomial distribution. Note that for the two-step method only the estimates of θ and σ_2^2 are included because estimated values of $\alpha_1^{\mathrm{CLS}}, \alpha_2^{\mathrm{CLS}}, \mu_{\varepsilon,1}^{\mathrm{CLS}}, \mu_{\varepsilon,2}^{\mathrm{CLS}}$ are used in order to estimate the remaining parameters via CML.

The results for the Poisson marginal distribution case are provided in Table 4.3.1. The results for the case when one innovation follows a Poisson distribution and the other – a negative binomial distribution are provided in Table 4.3.2. The lowest MSE values of $\widehat{\theta}$ are highlighted in bold. It is worth noting that CML estimation via numerical maximization depends heavily on the initial parameter values. If the initial values are selected too low or too high from the actual value, then the global maximum may not be found. In order to overcome this, we have selected the starting values equal to the CLS parameter estimates.

As can be seen in Table 4.3.1, the estimated values of α_j and $\mu_{\varepsilon,j}$, $j = 1,2$ have a smaller bias and MSE when parameters are estimated via CML. On the other hand, estimation of θ via CLS exhibits a smaller

Table 4.3.1: Monte Carlo simulation results for a BINAR(1) model with Poisson innovations linked by an FGM, Frank or Clayton copula.

Copula	Sample size	Param.	True value	CLS		CML		Two-Step	
				MSE	Bias	MSE	Bias	MSE	Bias
FGM	$N = 50$	α_1	0.6	0.01874	-0.05823	0.00887	-0.01789	-	-
		α_2	0.4	0.02033	-0.05223	0.01639	-0.02751	-	-
		$\mu_{\varepsilon,1}$	1	0.12983	0.13325	0.06514	0.03366	-	-
		$\mu_{\varepsilon,2}$	2	0.25625	0.16029	0.19939	0.07597	-	-
		θ	-0.5	**0.29789**	0.12568	0.33840	0.07568	0.3311	0.0876
	$N = 500$	α_1	0.6	0.00147	-0.00432	0.00073	-0.00122	-	-
		α_2	0.4	0.00184	-0.00505	0.00129	-0.00157	-	-
		$\mu_{\varepsilon,1}$	1	0.01012	0.00968	0.00556	0.00215	-	-
		$\mu_{\varepsilon,2}$	2	0.02413	0.01843	0.01763	0.00678	-	-
		θ	-0.5	0.04679	0.00668	0.04271	-0.00700	**0.04265**	-0.00443
Frank	$N = 50$	α_1	0.6	0.02023	-0.06039	0.00950	-0.01965	-	-
		α_2	0.4	0.02005	-0.05251	0.01630	-0.02858	-	-
		$\mu_{\varepsilon,1}$	1	0.13562	0.13536	0.06740	0.03625	-	-
		$\mu_{\varepsilon,2}$	2	0.25687	0.16392	0.19975	0.08291	-	-
		θ	-1	**1.83454**	0.12394	2.05786	0.00860	1.97515	0.04216
	$N = 500$	α_1	0.6	0.00153	-0.00595	0.00075	-0.00249	-	-
		α_2	0.4	0.00181	-0.00582	0.00129	-0.00132	-	-
		$\mu_{\varepsilon,1}$	1	0.01033	0.01269	0.00550	0.00421	-	-
		$\mu_{\varepsilon,2}$	2	0.02442	0.02129	0.01785	0.00629	-	-
		θ	-1	0.22084	0.01746	0.20138	-0.01779	**0.20070**	-0.01342
Clayton	$N = 50$	α_1	0.6	0.01826	-0.05489	0.00799	-0.013295	-	-
		α_2	0.4	0.01976	-0.05057	0.01585	-0.02427	-	-
		$\mu_{\varepsilon,1}$	1	0.12679	0.12104	0.06080	0.01743	-	-
		$\mu_{\varepsilon,2}$	2	0.25725	0.15704	0.19934	0.06499	-	-
		θ	1	0.71845	0.02621	0.72581	0.22628	**0.62372**	0.13283
	$N = 500$	α_1	0.6	0.00146	-0.00518	0.00070	0.00016	-	-
		α_2	0.4	0.00189	-0.00350	0.00120	-0.00049	-	-
		$\mu_{\varepsilon,1}$	1	0.00973	0.01137	0.00513	-0.00150	-	-
		$\mu_{\varepsilon,2}$	2	0.02447	0.01113	0.01707	0.00065	-	-
		θ	1	0.11578	0.03556	0.05864	0.04250	**0.03199**	-0.01342

MSE in the Frank copula case for smaller samples. For larger samples, the estimates of θ via the Two-step estimation method are very close to the CML estimates in terms of MSE and bias and are closer to the true parameter values, compared to the CLS estimates. Furthermore, since in the Two-step estimation numerical maximization is only carried out via a single parameter θ, the initial parameter values have less of an effect on the numerical maximization.

Table 4.3.2: Monte Carlo simulation results for a **BINAR**(1) model with one innovation following a Poisson distribution and the other – a negative binomial distribution, where both innovations are linked by an FGM, Frank or Clayton copula.

Copula	Sample size	Param.	True value	CLS		CML		Two-Step	
				MSE	Bias	MSE	Bias	MSE	Bias
FGM	$N=50$	α_1	0.6	0.01895	-0.05858	0.00845	-0.01513	-	-
		α_2	0.4	0.01936	-0.04902	0.00767	-0.01953	-	-
		$\mu_{\varepsilon,1}$	1	0.12940	0.12812	0.05424	0.01879	-	-
		$\mu_{\varepsilon,2}$	2	0.39724	0.15151	0.24138	0.04833	-	-
		θ	-0.5	0.31467	0.14070	**0.29415**	0.06674	0.29949	0.09693
		σ_2^2	9	27.87327	1.15731	15.12863	-0.14888	21.68229	0.72326
	$N=500$	α_1	0.6	0.00156	-0.00695	0.00076	-0.00153	-	-
		α_2	0.4	0.00194	-0.00373	0.00053	0.00016	-	-
		$\mu_{\varepsilon,1}$	1	0.01041	0.01201	0.00543	0.00290	-	-
		$\mu_{\varepsilon,2}$	2	0.03882	0.01843	0.02362	-0.00057	-	-
		θ	-0.5	0.06670	-0.02014	**0.04298**	-0.00268	0.04313	0.00562
		σ_2^2	9	6.24237	-1.99232	1.81265	0.00611	1.85222	-0.03506
Frank	$N=50$	α_1	0.6	0.02049	-0.06064	0.00912	-0.01594	-	-
		α_2	0.4	0.01951	-0.04936	0.00772	-0.02070	-	-
		$\mu_{\varepsilon,1}$	1	0.13769	0.13467	0.05748	0.02280	-	-
		$\mu_{\varepsilon,2}$	2	0.40626	0.15408	0.23717	0.05534	-	-
		θ	-1	1.81788	0.12516	1.75638	-0.01239	**1.68019**	0.06211
		σ_2^2	9	25.10400	0.49423	14.86812	-0.10034	21.92090	0.74026
	$N=500$	α_1	0.6	0.00161	-0.00702	0.00075	-0.00239	-	-
		α_2	0.4	0.00187	-0.00364	0.00050	-0.00046	-	-
		$\mu_{\varepsilon,1}$	1	0.01093	0.01652	0.00562	0.00501	-	-
		$\mu_{\varepsilon,2}$	2	0.03728	0.01217	0.02335	0.00203	-	-
		θ	-1	0.31942	-0.05593	**0.18960**	-0.01481	0.1902	-0.0079
		σ_2^2	9	4.82620	-1.75765	1.83082	0.02144	1.85852	-0.02690
Clayton	$N=50$	α_1	0.6	0.01987	-0.06159	0.00903	-0.01671	-	-
		α_2	0.4	0.01879	-0.04928	0.00632	-0.01644	-	-
		$\mu_{\varepsilon,1}$	1	0.13479	0.14072	0.06096	0.03052	-	-
		$\mu_{\varepsilon,2}$	2	0.40675	0.14807	0.23171	0.02871	-	-
		θ	1	0.78497	0.07464	0.67837	0.21235	**0.57454**	0.10972
		σ_2^2	9	24.40051	0.17321	15.29879	-0.08379	23.73506	0.73754
	$N=500$	α_1	0.6	0.00153	-0.00722	0.00075	-0.00197	-	-
		α_2	0.4	0.00196	-0.00385	0.00047	-0.00083	-	-
		$\mu_{\varepsilon,1}$	1	0.01036	0.01745	0.00517	0.00409	-	-
		$\mu_{\varepsilon,2}$	2	0.03999	0.01227	0.02304	0.00110	-	-
		θ	1	0.09927	0.04408	**0.05557**	0.03556	0.05559	0.02310
		σ_2^2	9	2.95995	-0.68733	1.79836	0.01348	1.87740	-0.02407

Table 4.3.2 demonstrates the estimation method results when one innovation follows a Poisson and the other has a negative binomial dis-

tribution.

With the inclusion of an additional variance parameter, CLS estimation methods exhibit larger MSE and bias for both the dependence and variance parameter estimates than the CML and Two-step estimation methods. Furthermore, the MSE of $\hat{\sigma}_2^2$ is the smallest when using CML estimation method. On the other hand, both Two-step and CML estimation methods produce similar estimates of θ in terms of MSE, regardless of sample size and copula function.

We can conclude that, while $\hat{\alpha}_j$, $\hat{\mu}_{\varepsilon,j}$ and $\hat{\sigma}_2^2$ are closer to the true parameter values via CML estimation method, it is possible to accurately estimate the dependence parameter via CML while using CLS estimates of $\hat{\alpha}_j$ and $\hat{\mu}_{\varepsilon,j}$. The resulting $\hat{\theta}$ will be closer to the actual value of θ compared to $\hat{\theta}^{\text{CLS}}$ and will not differ much from $\hat{\theta}^{\text{CML}}$. Additional inference on the bias of the estimates can be found in Appendix C.

4.4 Application to default loan data

In this section we estimate a $\mathsf{BINAR}(1)$ model with the joint innovation distribution modelled by a copula cdf for empirical data. The data set consists of loan data which includes loans that have defaulted and loans that were repaid without missing any payments. We will analyse and model the dependence between defaulted and non-defaulted loans as well as the presence of autocorrelation.

4.4.1 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 most older loans due to a lack of data needed for Bondoras rating model. Because Bondora issues loans in 4 different countries: Estonia, Finland, Slovakia and Spain, we will only focus on the loans issued in Spain. Since a new rating model indicates new rules for accepting or rejecting loans, we have selected the data sample from 21 October 2013, because from that date forward all loans had a rating assigned to them, to 1 January 2016. The time series are displayed in Figure 4.4.1. We are analysing data consisting of 115 weekly data.

- 'CompletedLoans' – the number of loans that were issued each week which are repaid and have never defaulted (a loan that is 60 or more days overdue is considered defaulted);
- 'DefaultedLoans' – the number of loans that were issued each week which have defaulted.

The loan statistics are provided in Table 4.4.1:

Table 4.4.1: Summary statistics of the weekly data of defaulted loans and non-defaulted loans issued in Spain.

	min	max	mean	variance
DefaultedLoans	1.00	60.00	22.60	158.66
CompletedLoans	0.00	15.00	5.30	11.67

The mean, minimum, maximum and variance is higher for defaulted loans compared to loans, which were repaid on time. As can be seen from Figure 4.4.2, the loans might be correlated since they both exhibit increase and decrease periods at the same times.

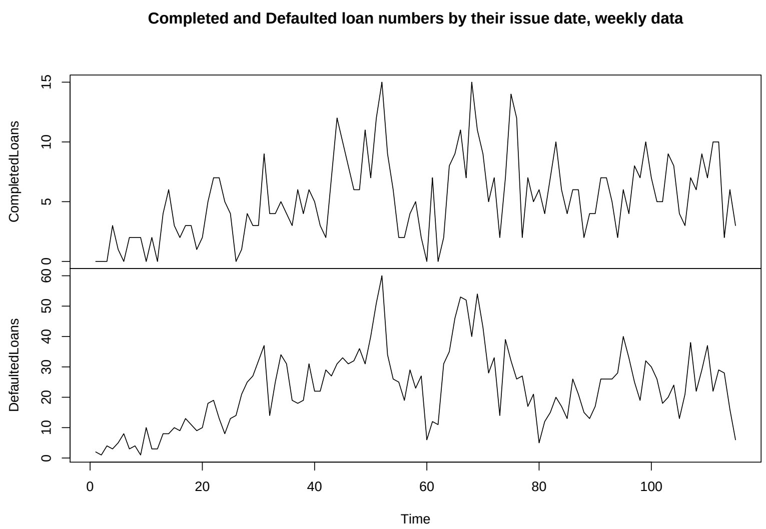


Figure 4.4.1: Bondora loan data: non-defaulted and defaulted loans by their issued date.

The correlation between the two time series is 0.6684. We also note that the mean and variance appears lower in the beginning of the time series. This 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 number of loans, issued to people living outside of Estonia, increased. The analysis of the significance of these effects is left for future research.

The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are displayed in Figure 4.4.2. We can see that the ACF function is decaying over time and the PACF function has a significant first lag which indicates that the non-negative integer-valued time series could be autocorrelated.

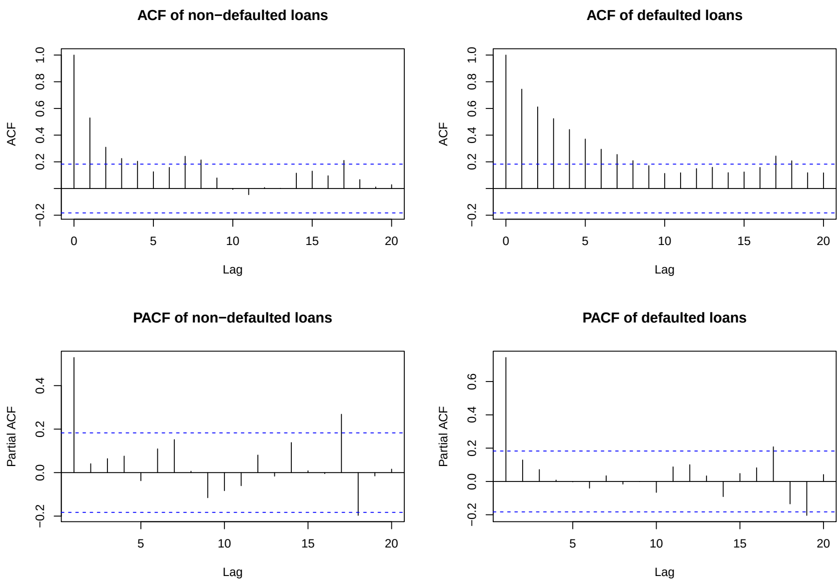


Figure 4.4.2: ACF and PACF plots of Bondora loan data.

We can also examine the heat map between the two series in Figure 4.4.3. We see that larger values of completed loans are associated with larger values of defaulted loans. However, when analysing BINAR(1) models, the heat map represents the relationship between the elements of $\mathbf{Y}_t$ in eq. (4.1.1). Since we do not observe the innovations - we cannot infer the underlying innovation copula as it is different from the distribution of the BINAR(1) process itself, see (Karlis and Pedeli, 2013).

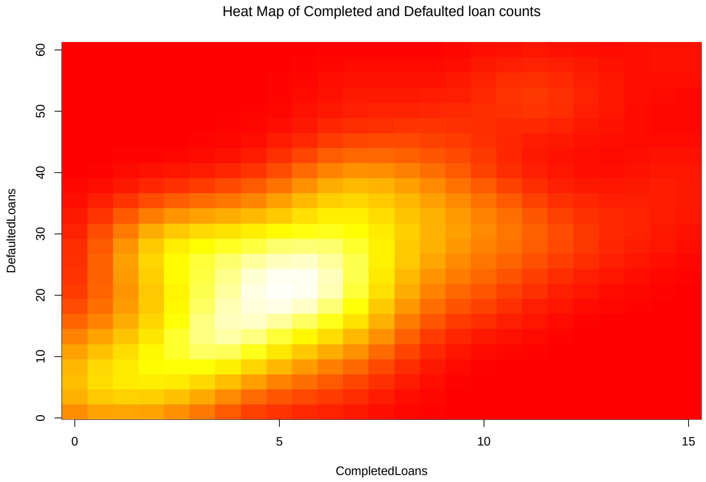


Figure 4.4.3: Heat map of completed and defaulted loan data.

Another well-known tool to visualize the cross-dependency between two time series is their cross-correlation plot (CCF), which is provided in Figure 4.4.4. We see that there is a relationship between the defaulted and completed loans. However, similarly to Figure 4.4.3, we cannot verify the underlying relationship of the innovations. An open question remains for selecting appropriate empirical methods for relating the distribution of the BINAR(1) process to the distribution of the unobserved innovations.

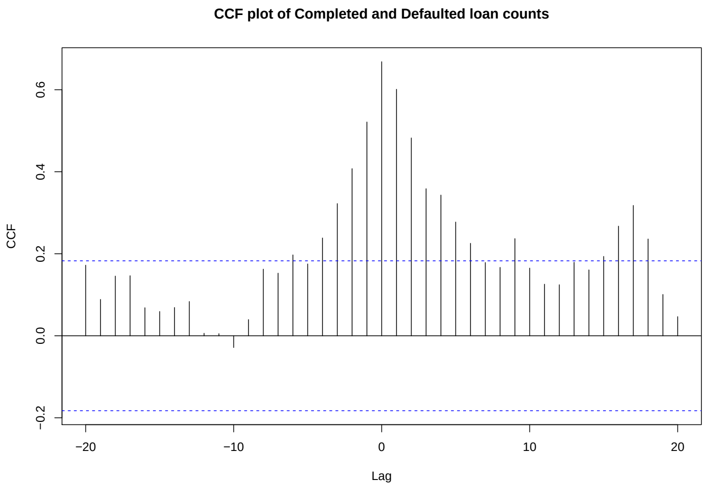


Figure 4.4.4: CCF plot of completed and defaulted loan data.

In order to analyse whether the number of loans that have defaulted depends on the number of loans which have not defaulted but were issued on the same week, 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. Based on the Monte Carlo simulation results presented in Section 4.3 and that our focus is the estimation of the dependence parameter, we will use the Two-step estimation method.

4.4.2 Estimated models

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

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

In the first step of the Two-step method, we estimated $\hat{\alpha}_1$ and $\hat{\mu}_{\varepsilon,1}$ for the non-defaulted loans, and $\hat{\alpha}_2$ and $\hat{\mu}_{\varepsilon,2}$ for the defaulted loans via CLS. The results are provided in Table 4.4.2 with standard errors for the Poisson case in parenthesis:

Table 4.4.2: Parameter estimates for BINAR(1) model via the Two-step estimation method: parameter CLS estimates from the first step with standard errors for the Poisson marginal distribution case in parenthesis.

$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_{\varepsilon,1}$	$\hat{\mu}_{\varepsilon,2}$
0.53134	0.75581	2.52174	5.58940
(0.08151)	(0.06163)	(0.45012)	(1.41490)

Because the CLS estimation method of parameters α_j and $\mu_{\varepsilon,j}$, $j = 1,2$ does not depend on the selected copula and marginal distribution family, these parameters will remain the same for each of the different distribution combinations for innovations. We can see that the defaulted loans exhibit a higher degree of autocorrelation, because of the larger value of $\hat{\alpha}_2$. The innovation mean parameter for defaulted loans is also higher in defaulted loans which indicates that random shocks have a larger effect on the number of defaulted loans.

The estimated parameter results from the second-step are provided in Table 4.4.3 with standard errors in parenthesis. $\hat{\sigma}_1^2$ is the innovation variance estimate of non-defaulted loans and $\hat{\sigma}_2^2$ is the innovation variance estimate of defaulted loans. According to Pawitan (2001), the observed Fisher information is the negative Hessian matrix, evaluated at the MLE. The asymptotic standard errors reported in Table 4.4.3 are derived under the assumption that α_j and $\mu_{\varepsilon,j},\ j=1,2$ are known, ignoring that the true values are substituted in the second step with their CLS estimates.

Table 4.4.3: Parameter estimates for **BINAR**(1) model via Two-step estimation method: parameter CML estimates from the second-step for different innovation marginal and joint distribution combinations with standard errors in parenthesis, derived under the assumption that the values $\hat{\mu}_{\varepsilon,j}$ and $\hat{\alpha}_j,\ j=1,2$ from the first step are true.

Marginals	Copula	$\hat{\theta}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	AIC	Log-likelihood
Both Poisson	FGM	0.89270 (0.18671)	-	-	1763.48096	-880.74048
	Frank	2.38484 (0.53367)	-	-	1760.15692	-879.07846
	Clayton	0.39357 (0.11697)	-	-	1761.12369	-879.56185
Negative binomial and Poisson	FGM	1.00000 (0.22914)	6.46907 (1.01114)	-	1731.57339	-863.78670
	Frank	2.14329 (0.45100)	6.10242 (1.15914)	-	1731.95241	-863.97620
	Clayton	0.34540 (0.12859)	5.73731 (0.52831)	-	1736.47641	-866.23821
Poisson and negative binomial	FGM	1.00000 (0.26357)	-	44.83107 (7.37423)	1498.29563	-747.14782
	Frank	2.01486 (0.61734)	-	44.10555 (7.33169)	1498.81039	-747.40519
	Clayton	0.38310 (0.17376)	-	43.42739 (7.29842)	1503.55388	-749.77694
Both negative binomial	FGM	1.00000 (0.31675)	6.55810 (1.24032)	45.36834 (7.55217)	**1466.15418**	**-730.07709**
	Frank	2.21356 (0.68192)	6.58754 (1.26126)	45.42601 (7.57743)	1466.97947	-730.48973
	Clayton	0.55939 (0.24652)	6.64478 (1.25833)	45.78307 (7.66324)	1470.73515	-732.36758

From the results in Table 4.4.3 we see that, according to the AIC and log-likelihood values, in most cases FGM copula most accurately describes the relationship between the innovations of defaulted and non-defaulted loans with Frank copula being very close in terms of AIC value. Clayton copula is the least accurate in describing the innovation joint distribution, when compared to FGM and Frank copula cases, which indicates that defaulted and non-defaulted loans do not exhibit strong left tail dependence.

Since the summary statistics of the data sample showed that the variance of the data is larger than the mean, a negative binomial marginal distribution may provide a better fit. Additionally, because copulas can link different marginal distributions, it is interesting to see if copulas with different discrete marginal distributions would also improve the model fit. BINAR(1) models where non-defaulted loan innovations are modelled with negative binomial and defaulted loan innovations are modelled with Poisson marginal distributions and vice versa were estimated. In general, changing one of the marginal distributions to a negative binomial provides a better fit in terms of AIC compared to the Poisson marginal distribution case. However, the smallest AIC value is achieved when both marginal distributions are modelled with negative binomial distributions, linked via an FGM copula. Furthermore, the estimated innovation variance is much larger for defaulted loans, $\hat{\sigma}_2^2$, which is similar to what we observed from the defaulted loan data summary statistics.

Overall, both Frank and FGM copulas provide similar fit in terms of log-likelihood, regardless of the selected marginal distributions. We note, however, that for some FGM copula cases, the estimated value of parameter θ is equal to the maximum attainable value of 1. Based on copula descriptions from Section 3, the FGM copula is used to model weak dependence. Given a larger sample size, a Frank copula might be more appropriate because it can capture a stronger dependence than that of an FGM copula. Looking at the negative binomial marginal distribution case $\hat{\theta} \approx 2.21356$ for the Frank copula, which indicates that there is a positive dependence between defaulted and non-defaulted loans just as in the FGM copula case.

4.5 Summary

The analysis via Monte Carlo simulations of different estimation methods shows that, although the estimates of BINAR(1) parameters via CML has the smallest MSE and bias, estimates of the dependence parameter has 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 parameter values. These values can be selected via CLS estimation.

An empirical application of BINAR models for loan data shows that, regardless of the selected marginal distributions, the FGM copula provides the best model fit in almost all cases with Frank copula being very close in terms of AIC values. For some of these cases, the estimated FGM copula dependence parameter value was equal to the maximum that can be attained by an FGM copula. In such cases, a larger sample size could help determine whether FGM or Frank copula is more appropriate to model the dependence between defaulted and non-defaulted loan amounts.

Although selecting marginal distributions from different families (Poisson or negative binomial) provided better models compared to models with only Poisson marginal distributions, the models with both marginal distributions modelled via negative binomial distributions provides the smallest AIC values which reflects overdispersion in both defaulted and non-defaulted loans. The FGM copula, which provides the best model fit, models variables, which exhibit weak dependence. Furthermore, the estimated copula dependence parameter indicates that the dependence between defaulted and non-defaulted loans is positive.

Finally, one can apply some other copulas in order to analyse whether the loan data exhibits different forms of dependence from the ones discussed in this paper. Lastly, the model can be extended by analysing the presence of structural changes within the data, or checking the presence of seasonality 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.

Chapter 5

An integer-valued autoregressive process for seasonality

In this chapter, we extend the $\text{INAR}(1)_d$ model in (1.0.4) by allowing the autoregressive parameter to vary with season as well as allowing the random shocks to be intra-seasonally dependent. In particular, we show that this specification leads to a similar form as the multivariate INAR(1) with dimension d. Also, we present the properties of the model and discuss the parameter estimation methods. Furthermore, a Monte Carlo simulation is carried out in order to compare parameter estimation methods. Finally, an empirical application is carried out on crime data in Chicago city.

5.1 $\text{SINAR}(1)_d$ process for seasonality

We propose a generalization of eq. (1.0.4) by allowing ϕ to vary based on the season as well as imposing a intra-seasonal dependence on the innovations. We define this INAR process as follows:

Definition 5.1.1. A sequence of nonnegative integer-valued random variables $\{Y_t, t \in \mathbb{Z}\}$ is said to be an INAR(1) process for seasonality with period $d \in \mathbb{N}$ ($\text{SINAR}(1)_d$), if it satisfies the following equation:

$$Y_t = \phi_t \circ Y_{t-d} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{5.1.1}$$

where $\phi_t = \alpha_j \in [0,1)$ for $t = j + kd$, $k \in \mathbb{Z}$ and $j \in \{1,\dots,d\}$, $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of intra-seasonally-dependent nonnegative integer-valued r.v.s. It is assumed that $\phi_t \circ Y_{t-d} = \sum_{i=1}^{Y_{j+(k-1)d}} B_{i,j,k-1}$, where $B_{i,j,k-1}$ are independent Bernoulli random variables, independent of $Y_{j+(k-1)d}$ and $\varepsilon_{j+(k-1)d}$ for all $k \in \mathbb{Z}$, $j \in \{1,\dots,d\}$, with $\mathbb{P}(B_{i,j,k-1} = 1) = \alpha_j$. It is also assumed that $[\varepsilon_{1+kd},\dots,\varepsilon_{d+kd}]^\top$ are independent of $[Y_{1+(k-s)d},\dots,Y_{d+(k-s)d}]^\top$ for $s \in \{1,2,...\}$ and that the distribution of $[\varepsilon_{1+kd},\dots,\varepsilon_{d+kd}]^\top$ can be described by a joint multivariate distribution or, alternatively, by a copula.

Furthermore, we will assume that the marginal distributions for ε_{j+kd}, $j \in \{1,\dots,d\}$ from Def. 5.1.1 are such that $\mathbb{E}(\varepsilon_{j+kd}) = \mu_{\varepsilon,j}$ and $\mathbb{V}\mathrm{ar}(\varepsilon_{j+kd}) = \sigma_{\varepsilon,jj} < \infty$ for all k, and ε_t are intra-seasonally-dependent with

$$\mathbb{C}\mathrm{ov}(\varepsilon_{i+kd}, \varepsilon_{j+ld}) = \begin{cases} \sigma_{\varepsilon,ij}, & k = l, \\ 0, & k \neq l, \end{cases}$$

i.e. vectors $[\varepsilon_{1+kd} - \mu_{\varepsilon,1},\dots,\varepsilon_{d+kd} - \mu_{\varepsilon,d}]^\top$ form a white noise sequence.

A larger value of α_j would indicate seasons where counts occur more frequently, while lower values would indicate seasonal periods with a lower frequency. Note that observations in (1.0.4) are driven by the first period values $Y_1,\dots,Y_d$, but the parameter ϕ indicates that the binomial variables have the same success probability in each seasonal period. By specifying the success probability to be different for each season in (5.1.1), we allow for larger (or smaller) counts in seasons, where they are more (or less) likely to occur, instead of averaging the success probability across the season. Allowing ε_t to be intra-seasonally-dependent, we account for possible dependence of the unobserved shocks within the period.

The following proposition summarizes the main properties of the defined $\mathrm{SINAR}(1)_d$ process. For any $j \in \{1,\dots,d\}$ denote the lattice $j + \mathbb{Z}d := \{j + kd, k \in \mathbb{Z}\}$.

Proposition 5.1.1. *Let $\{Y_t\}$ be a process in Def. 5.1.1 and let $t = j + kd$ with $j \in \{1, \ldots, d\}$, $k \in \mathbb{Z}$. Then:*

(a) the unique stationary marginal distribution of process $\{Y_t, t \in j + \mathbb{Z}d\}$ has the form

$$Y_t \overset{\mathrm{d}}{=} \sum_{l=0}^{\infty} \alpha_j^l \circ \varepsilon_{t-ld}; \tag{5.1.2}$$

(b) $\mathbb{E}(Y_t) = \dfrac{\mu_{\varepsilon,j}}{1-\alpha_j}$;

(c) $\mathbb{V}\mathrm{ar}(Y_t) = \dfrac{\sigma_{\varepsilon,jj} + \alpha_j \mu_{\varepsilon,j}}{1-\alpha_j^2}$;

(d) if $s \in \mathbb{Z}$, then

$$\mathbb{C}\mathrm{ov}(Y_t, Y_{t-s}) = \begin{cases} \alpha_j^{|m|} \dfrac{\sigma_{\varepsilon,jj} + \alpha_j \mu_{\varepsilon,j}}{1-\alpha_j^2}, & \textit{if } s = md,\ m \in \mathbb{Z}, \\ \dfrac{\sigma_{\varepsilon,ij}}{1-\alpha_i\alpha_j}, & \textit{if } s = j - i \neq 0 \textit{ with } i \in \{1, \ldots, d\}, \\ 0, & \textit{otherwise}; \end{cases}$$

(e) the conditional mean is $\mathbb{E}(Y_t | Y_{t-d}) = \phi_t Y_{t-d} + \mu_{\varepsilon,j}$.

The proofs of these properties are provided in Appendix D.1. We denote $\gamma(t,s) := \mathbb{C}\mathrm{ov}(Y_t, Y_{t-s})$ and $\rho(t,s) := \mathbb{C}\mathrm{orr}(Y_t, Y_{t-s})$.

Remark 1. It is assumed that the process defined in Def. 5.1.1 is such that either $\mathbb{E}(Y_{j+kd}) \neq \mathbb{E}(Y_{i+ld})$, or $\gamma(j+kd, s) \neq \gamma(i+ld, s)$ for some $i \neq j$ and $\forall k, l \in \mathbb{Z}$.

Following Hurd and Miamee (2007, Def. 1.4), we restate here that any process $\{X_t, t \in \mathbb{Z}\}$ with finite second moments is *periodically correlated* with some period $T > 0$ if T is the smallest value for which $\mathbb{E}(X_t) = \mathbb{E}(X_{t+T})$ and $\mathbb{C}\mathrm{ov}(X_t, X_s) = \mathbb{C}\mathrm{ov}(X_{t+T}, X_{s+T})$ for any $s, t \in \mathbb{Z}$. Taking into account Remark 1, we have that the series $\{Y_t\}$, defined in Def. 5.1.1, is periodically correlated with period d.

Periodically correlated sequences, as defined in Def. 5.1.1, are nonstationary. However, when the period is known, they are equivalent to a multivariate stationary processes, as discussed in Hurd and Miamee (2007). To better visualize this relationship, let

$j \in \{1,...,d\}$, $k \in \{1,...,n\}$, $Y_k^{(j)} := Y_{j+kd}$ and $\varepsilon_k^{(j)} := \varepsilon_{j+kd}$. Then (5.1.1) can be written as separate INAR(1) processes

$$Y_k^{(j)} = \alpha_j \circ Y_{k-1}^{(j)} + \varepsilon_k^{(j)},$$

where $\varepsilon_k^{(j)}$ and $\varepsilon_k^{(i)}$ are dependent r.v.s $\forall i, j \in \{1,...,d\}$. Consequently, we can obtain a multivariate INAR(1) process representation by blocking the univariate process, defined by eq. (5.1.1). This is also visualized in Table 5.1.1. We discuss this representation in the following section.

Table 5.1.1: A SINAR$(1)_d$ process, expressed as separate INAR(1) processes, where $Y_\bullet^{(j)}$ represents $Y_k^{(j)}, \forall k \in \{1,...,n\}$.

Period elements (e.g. months)	Period number (e.g. years)						
	1	2	...	k	...	n	
1							$Y_\bullet^{(1)}$
2							$Y_\bullet^{(2)}$
$\vdots$							$\vdots$
j				(j,k)			$Y_\bullet^{(j)}$
$\vdots$							$\vdots$
d							$Y_\bullet^{(d)}$

5.2 The multivariate representation of the SINAR$(1)_d$ process

Let $\{Y_t\}$ be a process defined in Def. 5.1.1. We can write the Y_t as a multivariate INAR(1) process

$$\mathbf{Y}_k = \mathbf{A} \circ \mathbf{Y}_{k-1} + \mathbf{Z}_k,\ k \in \mathbb{Z}, \tag{5.2.1}$$

where:

$$\mathbf{Y}_k = [Y_{1+kd}, \ldots, Y_{d+kd}]^\top,\ \mathbf{Z}_k = [\varepsilon_{1+kd}, \ldots, \varepsilon_{d+kd}]^\top,\ \mathbf{A} = \text{diag}(\alpha_1, \ldots, \alpha_d).$$

The properties of the binomial thinning operator are provided in Chapter 2.

Below we provide a number of properties of the multivariate INAR process, specified via eq. (5.2.1).

Proposition 5.2.1. *Let* $\{Y_t\}$ *be a solution of eq.* (5.1.1) *as in Proposition 5.1.1 (a). Then* $\{\mathbf{Y}_k\}$ *defined in eq.* (5.2.1) *is second-order stationary process and the following properties hold:*

(a) $\mathbb{E}(\mathbf{Y}_k) = \boldsymbol{\mu}_{\mathbf{Y}} := \left[\frac{\mu_{\varepsilon,1}}{1-\alpha_1}, \ldots, \frac{\mu_{\varepsilon,d}}{1-\alpha_d}\right]^\top = (\mathbf{I}-\mathbf{A})^{-1}\boldsymbol{\mu}_\varepsilon$,

where $\boldsymbol{\mu}_\varepsilon := \mathbb{E}(\mathbf{Z_k}) = [\mu_{\varepsilon,1}, \ldots, \mu_{\varepsilon,d}]^\top$;

(b) $\mathbb{E}(\mathbf{Y}_k|\mathbf{Y}_{k-1}) = \mathbf{A}\mathbf{Y}_{k-1} + \boldsymbol{\mu}_\varepsilon$;

(c) $\mathbb{C}\text{ov}(\mathbf{Z}_k, \mathbf{Z}_{k-h}) = \mathbb{E}\left(\mathbf{Z}_k - \mathbb{E}(\mathbf{Z}_k)\right)\left(\mathbf{Z}_{k-h} - \mathbb{E}(\mathbf{Z}_{k-h})\right)^\top = \begin{cases} \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}, & h = 0, \\ \mathbf{0}, & h > 0, \end{cases}$

where $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} := (\sigma_{\varepsilon,ij})_{i,j=1,\ldots,d}$;

(d) $\boldsymbol{\Gamma}(h) := \mathbb{C}\text{ov}(\mathbf{Y}_k, \mathbf{Y}_{k-h}) = \text{diag}(\alpha_1^h, \ldots, \alpha_d^h)\boldsymbol{\Gamma}(0),\ h \geq 0$,

where $\boldsymbol{\Gamma}(0) = \left(\frac{\sigma_{\varepsilon,ij} + \alpha_i\mu_{\varepsilon,i}\mathbb{1}_{\{i=j\}}}{1-\alpha_i\alpha_j}\right)_{i,j=1,\ldots,d}$.

The proofs of these properties are provided in Appendix D.2.

If $d = 2$, the multivariate specification in eq. (5.2.1) is equivalent to the bivariate INAR(1) model with copula-joint innovations, analysed in Chapter 4.

5.3 Parameter estimation of the SINAR$(1)_d$ process

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a sample drawn from a process satisfying eq. (5.2.1). In this section we will discuss some methods to estimate the unknown parameters $\alpha_j, \mu_{\varepsilon,j}, \sigma_{\varepsilon,jj},\ j = 1, \ldots, d$.

5.3.1 Parameter estimation via Restricted Estimated Generalized CLS

Rewrite $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ in the form of VAR(1) variables:

$$\mathbf{Y}_k = \boldsymbol{\mu}_\varepsilon + \mathbf{A}\mathbf{Y}_{k-1} + \boldsymbol{e}_k,\ k = 1, \ldots, n, \tag{5.3.1}$$

where $\boldsymbol{e}_k$ is the innovation vector

$$\boldsymbol{e}_k = \mathbf{Y}_k - \mathbb{E}(\mathbf{Y}_k|\mathbf{Y}_{k-1}) = \mathbf{A} \circ \mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1} + \mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon$$

with mean vector $\mathbb{E}(\boldsymbol{e}_k) = \mathbf{0}$ and covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{e}} = \operatorname{diag}(\mathbf{B}\boldsymbol{\mu}_{\mathbf{Y}}) + \boldsymbol{\Sigma}_{\varepsilon} = \left(\sigma_{\varepsilon,ij} + \alpha_i \mu_{\varepsilon,i} \mathbb{1}_{\{i=j\}}\right)_{i,j=1,\ldots,d}. \tag{5.3.2}$$

See Appendix D.3.

In order to estimate the parameters in (5.3.1), we introduce the following notations:

$$\begin{aligned}
&\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_n]^\top, \quad \mathbf{X}_k = [\mathbf{Y}_k^\top \vdots 1]^\top, \quad 0 \le k \le n-1, \\
&\mathbf{X} = [\mathbf{X}_0, \ldots, \mathbf{X}_{n-1}]^\top, \ \mathbf{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]^\top, \ \mathscr{B} = [\mathbf{A} \vdots \boldsymbol{\mu}_{\boldsymbol{\varepsilon}}]^\top.
\end{aligned}$$

Here, $\mathbf{Y}$ and $\mathbf{E}$ are $n \times d$ matrices, $\mathbf{X}_k$ is a $(d+1) \times 1$ vector, $\mathbf{X}$ is a $n \times (d+1)$ matrix, $\mathscr{B}$ is a $(d+1) \times d$ matrix.

Then we can write (5.3.1) as

$$\mathbf{Y} = \mathbf{X}\mathscr{B} + \mathbf{E}. \tag{5.3.3}$$

The conditional unrestricted least squares estimate of $\mathscr{B}$ is then:

$$\widetilde{\mathscr{B}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y}. \tag{5.3.4}$$

The asymptotic normality of the CLS estimator can be found in Latour (1997).

The multivariate INAR specification in eq. (5.2.1) restricts the coefficient matrix $\mathbf{A}$, so that non-diagonal elements are zero, while the coefficient estimate matrix in eq. (5.3.4) is unrestricted. To improve the efficiency of our estimates we impose coefficient restrictions on the non-diagonal elements. As in Lütkepohl (2007, Section 5.2), we impose zero-value constraints for $\mathscr{B}$ with the proposition below.

Let, further, '$\otimes$' and 'vec' respectively denote the Kronecker product and the vectorization of matrices.

Proposition 5.3.1. *Let* $\boldsymbol{\beta} := \operatorname{vec}(\mathscr{B}) = \mathbf{R}\boldsymbol{\gamma}$, *where* $\boldsymbol{\gamma} = [\alpha_1, \mu_{\varepsilon,1}, \ldots, \alpha_d, \mu_{\varepsilon,d}]^\top$ *is an unrestricted vector of unknown parameters and* $\mathbf{R}^\top = [\mathbf{R}_0^\top, \ldots, \mathbf{R}_{d-1}^\top]$ *is a known restriction matrix, where*

$\mathbf{R}_k = \left(r_{i,j}^{(k)}\right)$, *$k = 0,...,d-1$ are $(d+1)\times 2d$ matrices, with*

$$r_{i,j}^{(k)} = \begin{cases} 1, & \text{if either } i = k+1 \text{ and } j = 2k+1, \\ & \text{or } i = d+1 \text{ and } j = 2k+2, \\ 0, & \text{otherwise.} \end{cases} \tag{5.3.5}$$

Then the Restricted Estimated Generalized CLS (REG-CLS) estimator has the following form

$$\widehat{\boldsymbol{\beta}} = \mathbf{R}\widehat{\boldsymbol{\gamma}}, \quad \widehat{\boldsymbol{\gamma}} = \left[\mathbf{R}^\top\left(\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-1} \otimes \left(\mathbf{X}^\top\mathbf{X}\right)\right)\mathbf{R}\right]^{-1}\mathbf{R}^\top\left(\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-1} \otimes \mathbf{X}^\top\right)\mathrm{vec}\left(\mathbf{Y}\right), \tag{5.3.6}$$

where $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ is calculated as

$$\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = \frac{1}{n-d-1}(\mathbf{Y} - \mathbf{X}\widetilde{\mathscr{B}})^\top(\mathbf{Y} - \mathbf{X}\widetilde{\mathscr{B}}) \tag{5.3.7}$$

and $\widetilde{\mathscr{B}}$ is an unrestricted estimator from eq. (5.3.4)*.*

The proof is provided in Appendix D.4.

Example 5.3.1. Let $d = 3$, then $\boldsymbol{\beta} = [\alpha_1, 0, 0, \mu_{\varepsilon,1}, 0, \alpha_2, 0, \mu_{\varepsilon,2}, 0, 0, \alpha_3, \mu_{\varepsilon,3}]^\top$, $\gamma = [\alpha_1, \mu_{\varepsilon,1}, \alpha_2, \mu_{\varepsilon,2}, \alpha_3, \mu_{\varepsilon,3}]^\top$ and $\mathbf{R}^\top = [\mathbf{R}_0^\top \vdots \mathbf{R}_1^\top \vdots \mathbf{R}_2^\top]$. We can verify that with

$$\mathbf{R}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{R}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

$$\mathbf{R}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

it holds true that $\boldsymbol{\beta} = \mathbf{R}\boldsymbol{\gamma}$.

Proposition 5.3.2. *Let the process $\{\mathbf{Y}_k\}$ be defined in eq.* (5.2.1) *with independent white noise sequence $\{\mathbf{Z}_k\}$ with bounded fourth moments. Then the REG-CLS defined in eq.* (5.3.6) *is consistent and asymptotically* *normally distributed:*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\ \mathbf{R}\left[\mathbf{R}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}\otimes\boldsymbol{\Gamma}_{\boldsymbol{X}}(0)\right)\mathbf{R}\right]^{-1}\mathbf{R}^\top\right), \tag{5.3.8}$$

where $\dfrac{\mathbf{X}^\top\mathbf{X}}{n} \longrightarrow \boldsymbol{\Gamma}_{\boldsymbol{X}}(0) := \mathbb{E}\left(\mathbf{X}_k^\top\mathbf{X}_k\right)$ *in probability.*

The proof is presented in Appendix D.5.

In order to make inferences on the model parameters, we substitute $\boldsymbol{\Sigma}_{\boldsymbol{e}}$ with (5.3.7) and $\boldsymbol{\Gamma}_{\boldsymbol{X}}(0)$ with

$$\widehat{\boldsymbol{\Gamma}}_{\mathbf{X}}(0) = \frac{1}{n}\mathbf{X}^\top\mathbf{X}. \tag{5.3.9}$$

Furthermore, using (5.3.2) and (5.3.7), we can estimate $\Sigma_\varepsilon = (\sigma_{\varepsilon,ij})$, the covariance matrix of $\mathbf{Z}_k$, as

$$\widehat{\boldsymbol{\Sigma}}_\varepsilon = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} - \operatorname{diag}\left(\widehat{\alpha}_1\widehat{\mu}_{\varepsilon,1},\ldots,\widehat{\alpha}_d\widehat{\mu}_{\varepsilon,d}\right), \tag{5.3.10}$$

where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ is calculated using REG-CLS estimates as

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = (n-d-1)^{-1}(\mathbf{Y}-\mathbf{X}\widehat{\mathscr{B}})^\top(\mathbf{Y}-\mathbf{X}\widehat{\mathscr{B}}).$$

5.3.2 Parameter estimation via CML

We assume that the joint distribution of $[\varepsilon_{1+kd},\ldots,\varepsilon_{d+kd}]^\top$, denoted $F(\cdot)$, is described by a d-dimensional copula $C:[0,1]^d\to[0,1]$ with uniform margins, such that

$$F(a_1,\ldots,a_d) = C(F_1(a_1),\ldots,F_d(a_d)),$$

where $F_j(\cdot)$ is the univariate marginal distribution of ε_{j+kd}, $k\in\mathbb{Z}$. As outlined in Chapter 3, for discrete random variables, the copula is uniquely determined only on $\text{Range}(F_1)\times\ldots\times\text{Range}(F_d)$.

Following Nikoloulopoulos and Karlis (2009), the probability mass function of $[\varepsilon_{1+kd},...,\varepsilon_{d+kd}]$ is calculated as

$$
\begin{aligned}
c(a_1,\ldots,a_d;\boldsymbol{\theta}) &:= \mathbb{P}\left(\varepsilon_{1+kd}=a_1,...,\varepsilon_{d+kd}=a_d\right)\\
&= \sum_{i_1\in\{0,1\}}\cdots\sum_{i_d\in\{0,1\}}(-1)^{i_1+\ldots+i_d}\mathbb{P}\left(\varepsilon_{1+kd}\le a_1-i_1,...,\varepsilon_{d+kd}\le a_d-i_d\right)\\
&= \sum_{i_1\in\{0,1\}}\cdots\sum_{i_d\in\{0,1\}}(-1)^{i_1+\ldots+i_d}C(F_1(a_1-i_1),...,F_d(a_d-i_d);\boldsymbol{\theta}), \quad (5.3.11)
\end{aligned}
$$

where $\boldsymbol{\theta}=[\theta_1,...,\theta_m]^\top$ is the vector of m unknown copula dependence parameters and the marginal distribution of ε_{j+kd} has unknown mean $\mu_{\varepsilon,j}$ and variance $\sigma_{\varepsilon,jj}$.

Let the unknown parameter vector be

$$
\boldsymbol{\gamma} \;=\; [\alpha_1,\mu_{\varepsilon,1},\sigma_{\varepsilon,11},\ldots,\alpha_d,\mu_{\varepsilon,d},\sigma_{\varepsilon,dd},\theta_1,\ldots,\theta_m]^\top.
$$

Then, the conditional log likelihood function can be written as

$$
\ell(\boldsymbol{\gamma}|\mathbf{y}_0)=\sum_{k=1}^{n}\log(\mathbb{P}_{\mathbf{Y}_k|\mathbf{Y}_{k-1}}(\mathbf{y}_k|\mathbf{y}_{k-1})), \quad (5.3.12)
$$

for some initial values $\mathbf{y}_0=[y_1,\ldots,y_d]^\top$. Here $\mathbb{P}_{\mathbf{Y}_k|\mathbf{Y}_{k-1}}(\mathbf{y}_k|\mathbf{y}_{k-1})$ is the conditional probability of a process defined via eq. (5.2.1) with intra-seasonally-dependent shocks

$$
\begin{aligned}
\mathbb{P}_{\mathbf{Y}_k|\mathbf{Y}_{k-1}}(\mathbf{y}_k|\mathbf{y}_{k-1}) \;:=\; \mathbb{P}\left(\mathbf{Y}_k=\mathbf{y}_k|\mathbf{Y}_{k-1}=\mathbf{y}_{k-1}\right) \;=\; \mathbb{P}\left(\mathbf{A}\circ\mathbf{y}_{k-1}+\mathbf{Z}_k=\mathbf{y}_k\right)\\
=\sum_{j_1=0}^{y_{1+kd}}\cdots\sum_{j_d=0}^{y_{d+kd}}\mathbb{P}\left(\varepsilon_{1+kd}=j_1,\ldots,\varepsilon_{d+kd}=j_d\right)\prod_{i=1}^{d}p_i(j_i,y_{i+(k-1)d}),
\end{aligned}
$$

where $\mathbf{y}_k=[y_{1+kd},\ldots,y_{d+kd}]^\top$ and $p_i(j_i,y_{i+(k-1)d})$ is defined as the probability of the sum of $y_{i+(k-1)d}$ independent Bernoulli trials with j_i successes:

$$
\begin{aligned}
p_i(j_i,y_{i+(k-1)d}) &:= \mathbb{P}\left(\alpha_i\circ y_{i+(k-1)d}=y_{i+(k-1)d}-j_i\right)\\
&= \binom{y_{i+(k-1)d}}{y_{i+(k-1)d}-j_i}\alpha_i^{y_{i+(k-1)d}-j_i}(1-\alpha_i)^{j_i} \quad (5.3.13)
\end{aligned}
$$

with $j_i=0,\ldots,y_{i+kd}$, $i=1,...,d$, $k=1,...,n$. If the marginal distribution of the i-th residual ε_{i+kd} is Poisson with mean $\mu_{\varepsilon,i}$ ($\sigma_{\varepsilon,ii} = \mu_{\varepsilon,i}$), then

$$\mathbb{P}(\varepsilon_{i+kd} = j_i) = \mathrm{e}^{-\mu_{\varepsilon,i}} \frac{\mu_{\varepsilon,i}^{j_i}}{j_i!}. \tag{5.3.14}$$

If the marginal distribution of the i-th residual ε_{i+kd} is negative binomial with mean $\mu_{\varepsilon,i}$ and variance $\sigma_{\varepsilon,ii}$, then

$$\mathbb{P}\left(\varepsilon_{i+kd} = j_i\right) = \binom{j_i + \dfrac{\mu_{\varepsilon,i}}{\sigma_{\varepsilon,ii} - \mu_{\varepsilon,i}} - 1}{j_i} \left(\frac{\mu_{\varepsilon,i}}{\sigma_{\varepsilon,ii} - \mu_{\varepsilon,i}}\right)^{j_i} \left(\frac{\mu_{\varepsilon,i}}{\sigma_{\varepsilon,ii}}\right)^{\dfrac{\mu_{\varepsilon,i}^2}{\sigma_{\varepsilon,ii} - \mu_{\varepsilon,i}}}. \tag{5.3.15}$$

In order to estimate the unknown parameters, we maximize eq. (5.3.12):

$$\ell(\boldsymbol{\gamma}|\mathbf{y}_0) \longrightarrow \max_{\boldsymbol{\gamma}}. \tag{5.3.16}$$

Under some regularity conditions (see Franke and Seligmann (1993)), it can be shown that the CML estimate of $\boldsymbol{\gamma}$, obtained from (5.3.16) is asymptotically normal:

$$\sqrt{n}\left(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \xrightarrow{d} \mathscr{N}\left(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\gamma})\right), \tag{5.3.17}$$

where $\mathbf{I}(\boldsymbol{\gamma})$ is a $(3d+m) \times (3d+m)$ Fisher information matrix.

The drawback of estimating the model parameters by maximizing eq. (5.3.12) is that depending on the seasonality d, marginal distributions and copula function, the number of parameters to estimate may not be feasible in practical applications. To remedy this, we may use the inference function for margins (IFM) estimation method, which is a two-step likelihood-based estimation method.

5.3.3 Parameter estimation via IFM

Parameter estimation of a multivariate copula for count data is possible via IFM in two steps: (1) estimating parameters of the marginal distributions; (2) estimating the copula dependence parameters of the joint distribution using the marginal distribution parameters as fixed from the first step.

Let the unknown parameter vector of the marginal distributions of Y_{j+kd} be $\boldsymbol{\gamma}_j = [\alpha_j, \mu_{\varepsilon,j}, \sigma_{\varepsilon,jj}]^\top$, $j = 1, ..., d$. In the first step we independently maximize the conditional log-likelihood function for each marginal distribution

$$\ell_j(\boldsymbol{\gamma}_j | y_j) = \sum_{k=1}^{n} \log \left(\mathbb{P}_{Y_{j+kd}|Y_{j+(k-1)d}}(y_{j+kd}|y_{j+(k-1)d}) \right) \to \max_{\boldsymbol{\gamma}_j} \tag{5.3.18}$$

with some initial value y_j, $j = 1, ..., d$. Here, the conditional probability is given by

$$\begin{aligned} \mathbb{P}_{Y_{j+kd}|Y_{j+(k-1)d}}(y_{j+kd}|y_{j+(k-1)d}) &= \mathbb{P}(Y_{j+kd} = y_{j+kd}|Y_{j+(k-1)d} = y_{j+(k-1)d}) \\ &= \mathbb{P}(\alpha_j \circ y_{j+(k-1)d} + \varepsilon_{j+kd} = y_{j+kd}) \\ &= \sum_{i=0}^{y_{j+kd}} p_j(i, y_{j+(k-1)d}) \mathbb{P}(\varepsilon_{j+kd} = i), \end{aligned} \tag{5.3.19}$$

$j = 1, ..., d$, where $p_j(i, y_{j+(k-1)d})$ is defined in eq. (5.3.13) and $\mathbb{P}(\varepsilon_{j+kd} = i)$ is either (5.3.14) (in the case of Poisson marginal distribution) or (5.3.15) (in the case of negative binomial marginal distribution). In the second step, we estimate the copula parameter vector $\boldsymbol{\theta}$ by maximizing eq. (5.3.12) with the marginal distribution parameters fixed from eq. (5.3.18). For asymptotic efficiency and normality of the IFM, see Joe (2005).

5.4 Estimation method comparison via Monte Carlo simulation

In this section, we compare the performance of REG-CLS and IFM estimation methods by a Monte Carlo simulation study. We have simulated 5000 samples for the case $d = 2$ and 1000 samples for the case $d = 4$. The resulting mean squared errors (MSE) are presented in Tables 5.4.1, 5.4.2 and 5.4.3, where the Pois rows indicate the cases where the innovation marginal distributions are all Poisson; the NB rows indicate the cases where the innovation marginal distributions are all negative binomial; the Pois-NB rows indicate the cases where the innovation marginal distributions alternate between Poisson and negative binomial. In the case $d = 2$, the marginal distribution of ε_{1+2k} follows a Poisson distribution with $\mu_{\varepsilon,1}$ and the marginal distribution of ε_{2+2k} follows a negative binomial distribution with parameters $\mu_{\varepsilon,2}$ and $\sigma_{\varepsilon,22}$. For $d = 4$ the marginal distributions are $\varepsilon_{1+2k} \sim$ Pois, $\varepsilon_{2+2k} \sim$ NegBin, $\varepsilon_{3+2k} \sim$ Pois and $\varepsilon_{4+2k} \sim$ NegBin with their respective mean and variance parameters, which are indicated in Table 5.4.1, Table 5.4.2 and Table 5.4.3.

Table 5.4.1: MSE of the Monte Carlo simulation results for 5000 simulations of a $\text{SINAR}(1)_\text{d}$ model with different innovation marginal and copula distributions.

$d = 2$		α_1	α_2	$\mu_{\varepsilon,1}$	$\mu_{\varepsilon,2}$	$\sigma_{\varepsilon,11}$	$\sigma_{\varepsilon,22}$	θ
				Clayton				
True values	(Pois)	0.76	0.28	1	2	1	2	5
$N = 240$	REG-CLS	0.0040	0.0071	0.0681	0.0670	0.0742	0.1392	-
	IFM	0.0013	0.0074	0.0232	0.0694	-	-	20.3640
$N = 540$	REG-CLS	0.0015	0.0029	0.0267	0.0282	0.0318	0.0598	-
	IFM	0.0006	0.0031	0.0102	0.0294	-	-	7.6050
True values	(NB)	0.76	0.28	1	2	2.25	4.50	5
$N = 240$	REG-CLS	0.0038	0.0067	0.0784	0.0884	0.5033	1.1514	-
	IFM	0.0011	0.0054	0.0325	0.0783	0.4533	1.0487	10.0654
$N = 540$	REG-CLS	0.0016	0.003	0.0324	0.0374	0.2200	0.4876	-
	IFM	0.0005	0.0023	0.0137	0.0318	0.1990	0.4503	4.1426
True values	(Pois-NB)	0.76	0.28	1	2	1	4.50	5
$N = 240$	REG-CLS	0.0039	0.0071	0.0663	0.0899	0.0721	1.1433	-
	IFM	0.0013	0.0053	0.0232	0.0749	-	1.0527	12.3026
$N = 540$	REG-CLS	0.0015	0.0031	0.0264	0.0389	0.0308	0.5061	-
	IFM	0.0005	0.0022	0.0098	0.0326	-	0.4627	3.8763
				Frank				
True values	(Pois)	0.76	0.28	1	2	1	2	10
$N = 240$	REG-CLS	0.0039	0.0067	0.0691	0.0655	0.0743	0.1422	-
	IFM	0.0013	0.0074	0.0238	0.0695	-	-	15.7747
$N = 540$	REG-CLS	0.0016	0.0030	0.0278	0.0290	0.0326	0.0626	-
	IFM	0.0006	0.0032	0.0106	0.0301	-	-	6.9576
True values	(NB)	0.76	0.28	1	2	2.25	4.50	10
$N = 240$	REG-CLS	0.0039	0.0067	0.0793	0.0885	0.5205	1.1454	-
	IFM	0.0011	0.0054	0.0314	0.0758	0.4804	1.0601	8.3183
$N = 540$	REG-CLS	0.0015	0.0031	0.0322	0.0378	0.2213	0.5084	-
	IFM	0.0005	0.0023	0.0135	0.0317	0.1978	0.4655	2.9985
True values	(Pois-NB)	0.76	0.28	1	2	1	4.50	10
$N = 240$	REG-CLS	0.0037	0.0070	0.0616	0.0870	0.0719	1.1266	-
	IFM	0.0013	0.0051	0.0221	0.0722	-	1.0427	12.2461
$N = 540$	REG-CLS	0.0015	0.0031	0.0272	0.0395	0.0323	0.4879	-
	IFM	0.0006	0.0023	0.0105	0.0316	-	0.4507	4.5451

The IFM estimation method is carried out under the assumption of correctly specified margins and copula function. As a result, the estimates of the likelihood-based approach are more accurate and the variance for the Poisson marginal distributions is not estimated. On the other hand, the REG-CLS method uses eq. (5.3.10) to estimate the innovation covariance matrix and does not make any underlying assumptions about their joint and marginal distributions. Consequently, REG-CLS results for the Poisson marginal distributions in Pois and Pois-NB cases include the estimates for the variance parameter as well. These results indicate that estimated variance parameters are close to the mean estimates. This also confirms a well-known fact that the variance estimate of a Poisson r.v. has a larger MSE than the mean estimate. For negative binomial marginal distributions the larger variance results in less accurate variance estimates and mean estimates for the NB and Pois-NB cases. On the other hand, the accuracy of mean estimates for the Poisson marginal distributions for the Pois-NB cases is similar to the Pois cases. The estimates of α_j, $j = 1, ..., 4$ are unaffected neither by the marginal distribution of the innovations, nor by the underlying copula.

Furthermore, the density estimates of the parameters from the Monte Carlo simulation are presented in Fig. 5.4.1, which shows that increasing the sample size results in more accurate estimates for both REG-CLS and IFM estimation methods. This means that for larger samples, the REG-CLS estimates of the mean and variance will be close to the true values, regardless of the underlying marginal distribution. However, unlike the IFM estimation method, the REG-CLS estimation method does not allow to estimate the dependence parameter.

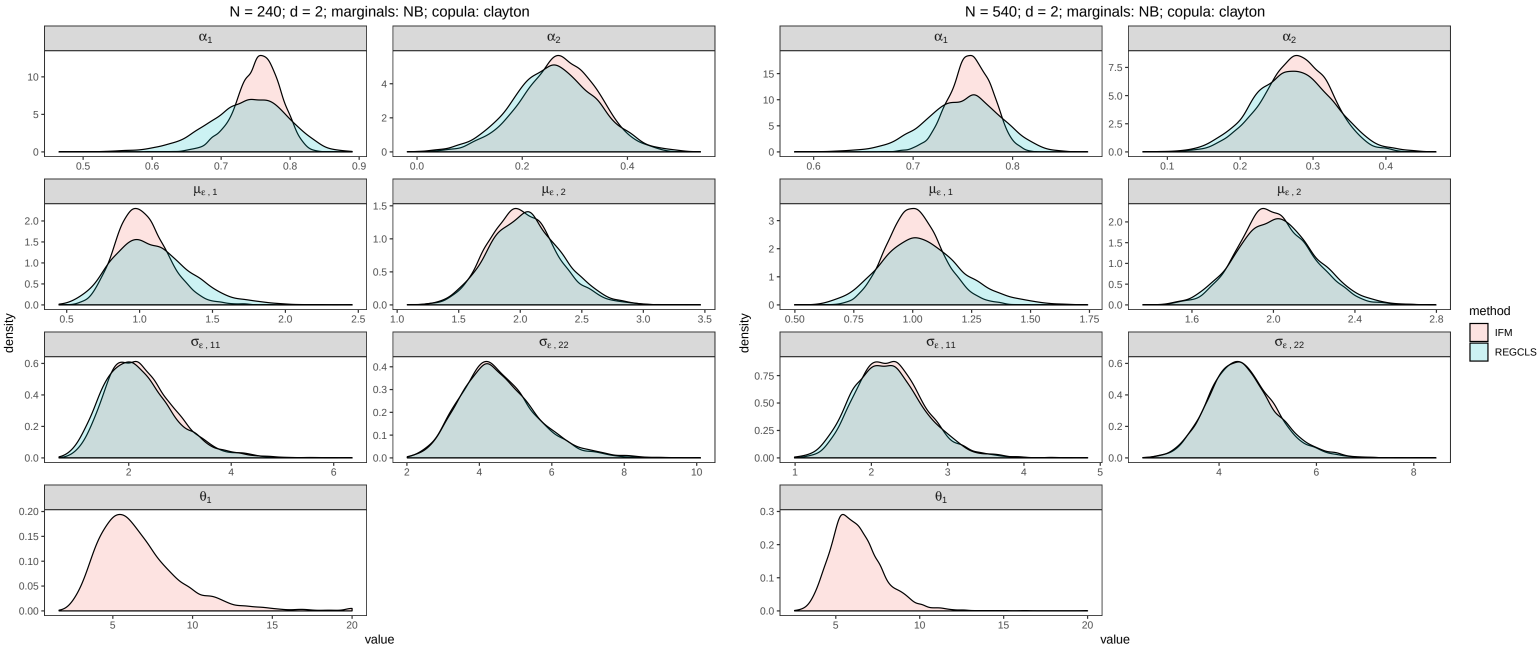
Finally, we would also like to mention that the overall calculation time of REG-CLS is much shorter than that of the IFM, especially for $d > 2$ since the copula pmf in eq. (5.3.11) requires 2^d copula evaluations for any given value vector $[a_1, ..., a_d]$. This limitation of the IFM for the discrete copula case is also discussed in Panagiotelis et al. (2012), where a discrete analogue to vine Pair Copula Constructions (PCC's) is proposed. The applicability of PCC's to $\mathrm{SINAR}(1)_\mathrm{d}$ is left for future research.

Table 5.4.2: MSE of the Monte Carlo simulation results for 1000 simulations of a $\mathbf{SINAR}(1)_{\mathrm{d}}$ model with different innovation marginal distributions and Clayton copula.

$d=4$		α_1	α_2	α_3	α_4	$\mu_{\varepsilon,1}$	$\mu_{\varepsilon,2}$	$\mu_{\varepsilon,3}$	$\mu_{\varepsilon,4}$	$\sigma_{\varepsilon,11}$	$\sigma_{\varepsilon,22}$	$\sigma_{\varepsilon,33}$	$\sigma_{\varepsilon,44}$	θ
							Clayton							
True values	(Pois)	0.76	0.9	0.72	0.28	1	2	3	4	1	2	3	4	5
$N=240$	REG-CLS	0.0094	0.0019	0.0077	0.0117	0.1588	0.5302	0.8571	0.4162	0.1576	0.8386	1.2116	1.2616	-
	IFM	0.0029	0.0004	0.0034	0.0156	0.0492	0.1302	0.3738	0.5269	-	-	-	-	17.1589
$N=540$	REG-CLS	0.0032	0.0009	0.0032	0.0046	0.0533	0.3010	0.3650	0.1649	0.0619	0.4276	0.4734	0.5228	-
	IFM	0.0012	0.0002	0.0012	0.0064	0.0195	0.0683	0.1354	0.2222	-	-	-	-	10.0703
True values	(NB)	0.76	0.9	0.72	0.28	1	2	3	4	2.25	4.5	6.75	9	5
$N=240$	REG-CLS	0.0084	0.0030	0.0072	0.0099	0.1799	0.9662	0.8815	0.4813	0.9046	2.8677	5.9890	7.6571	-
	IFM	0.0029	0.0008	0.0058	0.0126	0.0769	0.3171	0.7117	0.5227	0.7940	2.3899	5.0700	5.8236	9.2140
$N=540$	REG-CLS	0.0032	0.0011	0.0025	0.0043	0.0630	0.3988	0.3203	0.1866	0.4441	1.3379	2.1321	2.8840	-
	IFM	0.0009	0.0003	0.0018	0.0058	0.0278	0.1081	0.2302	0.2431	0.4035	1.1014	1.9504	2.5655	7.7525
True values	(Pois-NB)	0.76	0.9	0.72	0.28	1	2	3	4	1	4.5	3	9	5
$N=240$	REG-CLS	0.0095	0.0027	0.0076	0.0112	0.1436	0.8255	0.8256	0.4728	0.1475	3.4099	1.0535	7.3995	-
	IFM	0.0029	0.0007	0.0031	0.0125	0.0412	0.2667	0.3185	0.5301	-	2.5708	-	6.0270	10.7848
$N=540$	REG-CLS	0.0031	0.0011	0.0031	0.0045	0.0546	0.4252	0.3492	0.2043	0.0668	1.4936	0.5016	3.1498	-
	IFM	0.0011	0.0003	0.0014	0.0055	0.0211	0.1046	0.1667	0.2401	-	1.1617	-	2.7568	8.3908

Table 5.4.3: MSE of the Monte Carlo simulation results for 1000 simulations of a **SINAR**(1)$_{\mathrm{d}}$ model with different innovation marginal distributions and Frank copula.

$d = 4$		α_1	α_2	α_3	α_4	$\mu_{\varepsilon,1}$	$\mu_{\varepsilon,2}$	$\mu_{\varepsilon,3}$	$\mu_{\varepsilon,4}$	$\sigma_{\varepsilon,11}$	$\sigma_{\varepsilon,22}$	$\sigma_{\varepsilon,33}$	$\sigma_{\varepsilon,44}$	θ
							Frank							
True values	(Pois)	0.76	0.9	0.72	0.28	1	2	3	4	1	2	3	4	10
$N = 240$	REG-CLS	0.0095	0.0020	0.0080	0.0120	0.1557	0.5570	0.8351	0.4307	0.1498	0.8469	1.0703	1.2883	-
	IFM	0.0031	0.0005	0.0030	0.0155	0.0460	0.1473	0.3370	0.5312	-	-	-	-	12.7514
$N = 540$	REG-CLS	0.0033	0.0010	0.0032	0.0049	0.0549	0.3208	0.3582	0.1754	0.0599	0.4214	0.5279	0.5030	-
	IFM	0.0012	0.0002	0.0014	0.0067	0.0197	0.0646	0.1602	0.2277	-	-	-	-	5.0582
True values	(NB)	0.76	0.9	0.72	0.28	1	2	3	4	2.25	4.5	6.75	9	10
$N = 240$	REG-CLS	0.0083	0.0028	0.0071	0.0111	0.1692	0.8597	0.8598	0.4702	1.0231	2.8696	4.9963	6.9865	-
	IFM	0.0028	0.0007	0.0052	0.0133	0.0714	0.2876	0.6809	0.5654	0.8969	2.3213	4.3173	5.8163	6.9854
$N = 540$	REG-CLS	0.0029	0.0010	0.0026	0.0040	0.0653	0.3636	0.3244	0.1921	0.4492	1.3239	1.9567	3.0079	-
	IFM	0.0010	0.0002	0.0016	0.0051	0.0276	0.1099	0.2275	0.2314	0.4077	1.0185	1.7941	2.7375	2.5738
True values	(Pois-NB)	0.76	0.9	0.72	0.28	1	2	3	4	1	4.5	3	9	10
$N = 240$	REG-CLS	0.0096	0.0030	0.0073	0.0099	0.1395	0.9140	0.7660	0.4346	0.1492	3.6768	1.1012	7.7152	-
	IFM	0.0032	0.0008	0.0031	0.0122	0.0471	0.2743	0.3411	0.5158	-	2.8485	-	5.9642	7.3902
$N = 540$	REG-CLS	0.0034	0.0011	0.0028	0.0045	0.0603	0.4156	0.3160	0.1971	0.0639	1.3612	0.4817	2.8819	-
	IFM	0.0012	0.0002	0.0013	0.0054	0.0219	0.1025	0.1490	0.2167	-	1.0231	-	2.4537	3.1838



(a) $N = 240$, $d = 2$ with NB marginals and Clayton copula

(b) $N = 540$, $d = 2$ with NB marginals and Clayton copula

Figure 5.4.1: Monte Carlo simulation kernel density plots of the parameter estimates via IFM and REG-CLS estimation methods.

5.5 Empirical application to Chicago crime data

In this section, we carry out an empirical application using Chicago crime data[1] from January 2016 to March 2018. The data set includes the numbers of various types of crimes committed throughout the day. We have selected four types of crimes – assault, battery, burglary and robbery – and aggregated the data every 4 hours so as to divide each day into 6 periods. Their summary statistics are provided in Table 5.5.1.

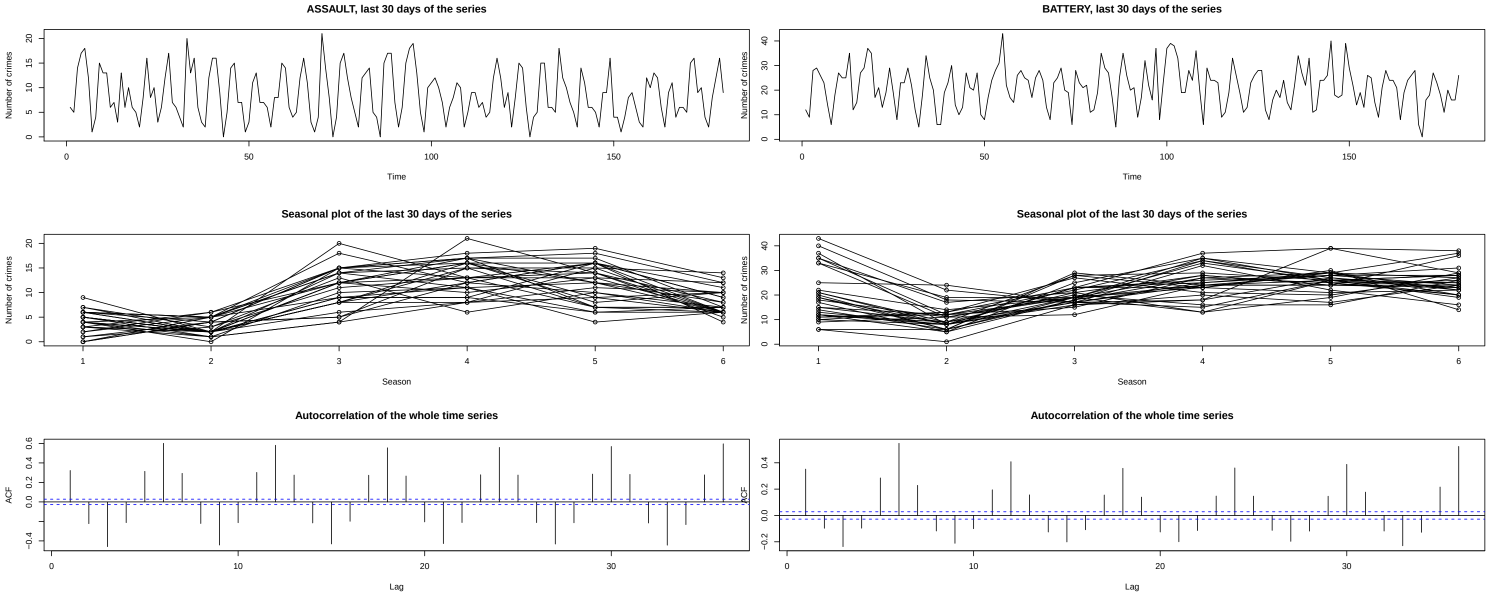
Table 5.5.1: Chicago Crime data summary statistics by crime type, aggregated by 4-hour periods.

Period (hour interval)	Mean	Variance	Minimum	Maximum
Assault				
00:00 - 03:59	4.4915	7.3955	0	17
04:00 - 07:59	3.1890	3.6602	0	12
08:00 - 11:59	9.7902	13.9853	1	22
12:00 - 15:59	12.5915	16.5496	2	28
16:00 - 19:59	12.2683	16.3699	2	26
20:00 - 23:59	9.2634	13.7938	1	24
Battery				
00:00 - 03:59	22.3707	142.5046	4	93
04:00 - 07:59	11.1366	28.5650	0	40
08:00 - 11:59	19.8195	25.8087	5	40
12:00 - 15:59	24.9183	35.4817	8	49
16:00 - 19:59	27.4024	45.1956	10	52
20:00 - 23:59	28.5671	75.5547	7	70
Burglary				
00:00 - 03:59	4.1427	5.5095	0	13
04:00 - 07:59	4.7159	6.7897	0	14
08:00 - 11:59	7.9963	13.9817	0	24
12:00 - 15:59	7.3280	9.5003	1	20
16:00 - 19:59	6.7780	9.3231	0	19
20:00 - 23:59	5.2134	6.7542	0	15
Robbery				
00:00 - 03:59	5.5598	10.7791	0	21
04:00 - 07:59	3.2024	5.0408	0	14
08:00 - 11:59	3.5683	4.4605	0	14
12:00 - 15:59	5.1000	5.4772	0	14
16:00 - 19:59	6.7268	10.9094	0	20
20:00 - 23:59	7.6659	10.3839	0	20

[1] for the latest data snapshot, see `https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2`

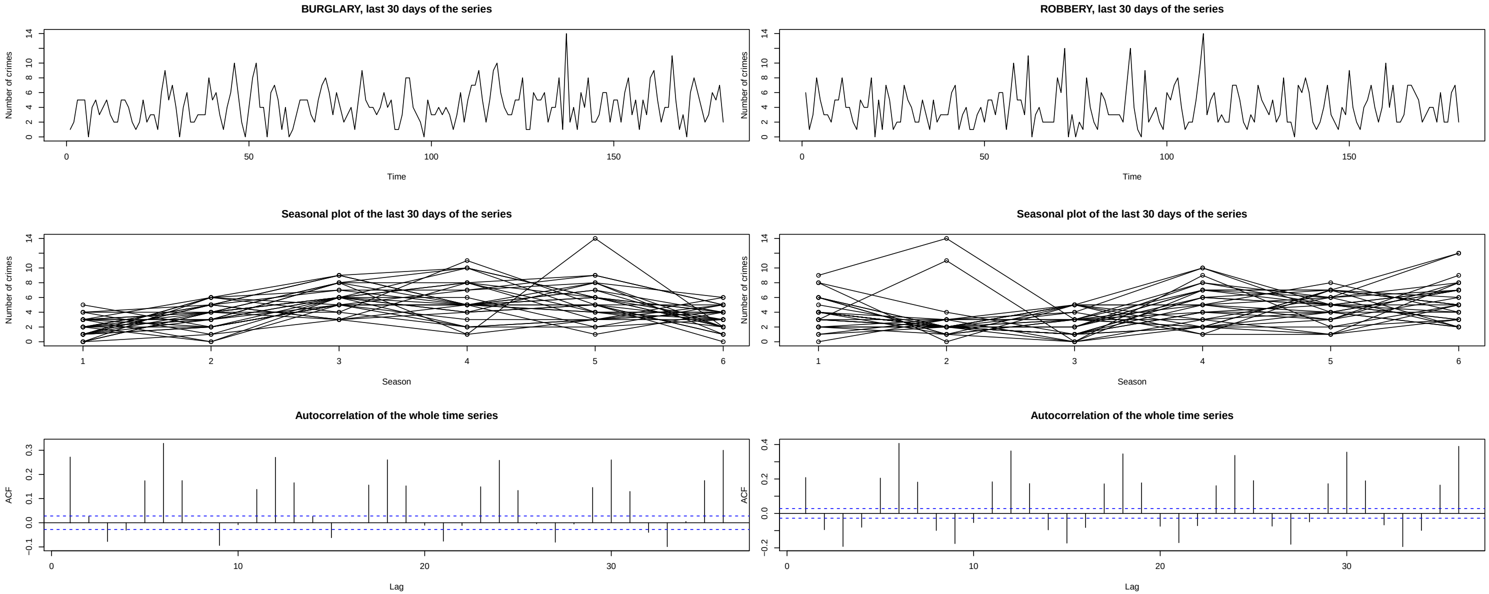
We see that the average number of offences occur at night time with battery having the largest variance. We also note that burglary and robbery have much lower number of occurrences compared to assault and battery.

The four different crime types exhibit different magnitudes in the mean and variance but they generally exhibit a similar autocorrelation pattern, as can be seen in Figure 5.5.1 and Figure 5.5.2.



(a) Assault **(b)** Battery

Figure 5.5.1: Chicago Crime data time series, aggregated by 4-hour periods, and seasonal plots for the last 30 day period from 2018-03-01 00:00 to 2018-03-30 23:59 and the full series autocorrelation plots for assault (left) and battery (right).



(a) Burglary

(b) Robbery

Figure 5.5.2: Chicago Crime data time series, aggregated by 4-hour periods, and seasonal plots for the last 30 day period from 2018-03-01 00:00 to 2018-03-30 23:59 and the full series autocorrelation plots for burglary (left) and robbery (right).

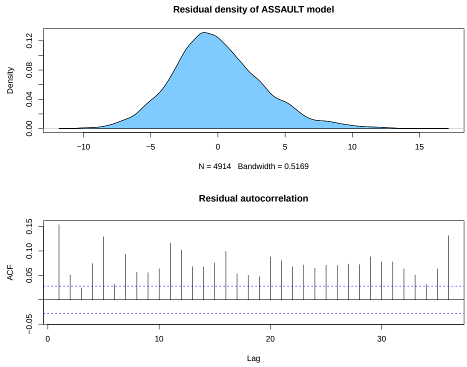
In order to capture the seasonality, we have fitted a SINAR$(1)_6$ model for each crime type and estimated the parameters via REG-CLS. The estimates of parameters are provided in Table 5.5.2.

Table 5.5.2: Estimated coefficients of the SINAR$(1)_6$ models on each crime type.

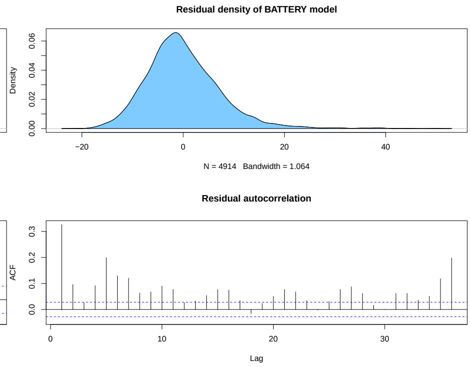
	Assault	Battery	Burglary	Robbery
α_1	0.1346	0.2164	0.1807	0.2897
	(0.0326)	(0.0243)	(0.0328)	(0.0317)
α_2	0.0712	0.1235	0.1462	0.0507
	(0.1725)	(0.6478)	(0.1568)	(0.2059)
α_3	0.1372	0.1026	0.1270	0.1149
	(0.0343)	(0.0278)	(0.0318)	(0.0334)
α_4	0.1272	0.1113	0.0947	0.1061
	(0.1276)	(0.3544	(0.1737)	(0.1318)
α_5	0.1081	0.2229	0.1035	0.3030
	(0.0321)	(0.0330)	(0.0308)	(0.0340)
α_6	0.1401	0.1391	0.1183	0.1309
	(0.3385)	(0.6763)	(0.2772)	(0.1416)
$\mu_{\varepsilon,1}$	3.8742	17.4434	3.3927	3.9476
	(0.0327)	(0.0327)	(0.0326)	(0.0341)
$\mu_{\varepsilon,2}$	2.9622	9.7319	4.0282	3.0414
	(0.4343)	(0.8393)	(0.2613)	(0.1916)
$\mu_{\varepsilon,3}$	8.4522	17.7855	6.9892	3.1601
	(0.0331)	(0.0306)	(0.0325)	(0.0326)
$\mu_{\varepsilon,4}$	10.9990	22.1420	6.6365	4.5616
	(0.4279)	(0.8672)	(0.2426)	(0.2449)
$\mu_{\varepsilon,5}$	10.9500	21.2918	6.0850	4.6894
	(0.0329)	(0.0298)	(0.0333)	(0.0335)
$\mu_{\varepsilon,6}$	7.9670	24.5938	4.6028	6.6662
	(0.3287)	(0.8985)	(0.1949)	(0.2795)
$\sigma_{\varepsilon,11}$	6.5824	119.3875	4.7119	8.5867
$\sigma_{\varepsilon,22}$	3.4522	26.0592	6.0079	4.8942
$\sigma_{\varepsilon,33}$	12.5246	23.7519	12.5929	4.0556
$\sigma_{\varepsilon,44}$	14.7102	32.3107	8.7780	4.9609
$\sigma_{\varepsilon,55}$	14.8733	37.0851	8.5753	8.4566
$\sigma_{\varepsilon,66}$	12.2922	67.0528	6.0547	9.3179
MSE	11.5855	53.4522	8.3737	7.3971

The estimated innovation mean and variance parameters are close to the sample values in Table 5.5.1. In order to assess the models, in Table 5.5.2 we also provide the mean squared errors (MSE). Since battery-type crime counts have the largest variance, we see that the latter model has the largest MSE among other types of crime models. The residual density and autocorrelation plots are shown in Figure 5.5.3, which indicate that seasonality was captured quite well by the specified $\mathrm{SINAR}(1)_6$ models. However, as indicated by the residual autocorrelation plots, the non-seasonal autocorrelation is not captured by the model.

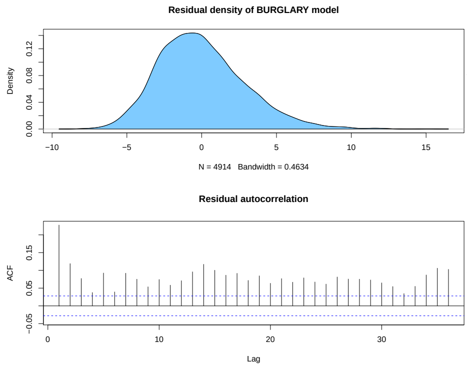
Finally, in order to capture the remaining significant non-seasonal autocorrelation, the $\mathrm{SINAR}(1)_6$ model could be extended to allow a non-seasonal autocorrelation component. Analysis of such a process is left for future research.



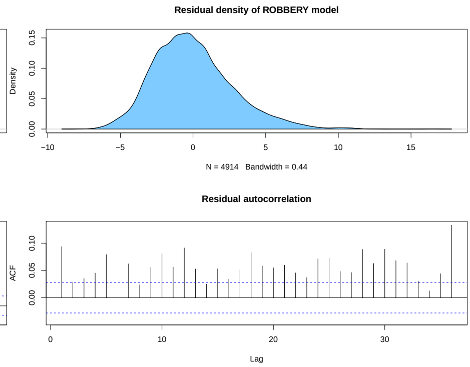
(a) Assault



(b) Battery



(c) Burglary



(d) Robbery

Figure 5.5.3: Residual density and autocorrelation plots of the fitted $\mathrm{SINAR}(1)_6$ models.

5.6 Summary

In this chapter a first order INAR process with seasonally-varying autocorrelation parameters and intra-seasonally-dependent shocks was proposed. Model properties for both the univariate and the multivariate representation of the model were provided with the multivariate form being stationary. In order to estimate the model parameters of the multivariate specification, a restricted estimated generalized conditional least squares (REG-CLS) estimation method with restrictions on the non-diagonal model coefficients was proposed.

A Monte Carlo simulation experiment was carried out in order to compare the aforementioned estimation method with the inference function of margins (IFM) method, which is a likelihood-based parameter estimation method. The simulation results showed that the REG-CLS method is much faster than the IFM method and produces similar results in terms of parameter accuracy for larger samples, although it does not allow estimation of the innovation copula dependence parameters. On the other hand, the IFM estimation method is computationally intensive, with the number of evaluations exponentially increasing with the seasonal period size.

Finally, an empirical application to four types of crime data in Chicago city was carried out. The estimated models indicated that the $\text{INAR}(1)_6$ can successfully be used to capture the seasonality for Chicago crime occurrence data time series. The empirical application results also indicated that the model could be extended in order to capture the non-seasonal autocorrelation effect. This is the topic of the authors' current research.

Conclusions

The main objective of this thesis is to provide some contributions to the analysis of integer-valued time series models. The main focus is on such models, where the joint innovation distribution can be described by a copula.

Firstly, a class of bivariate integer-valued autoregressive processes of order 1 (BINAR(1)) are analysed. Such a process can be thought of as a multivariate representation of two time series with copula-joint innovations. The existing literature on such models is extended by providing a two-step estimation method, where the BINAR(1) model parameters are estimated separately from the dependence parameter of the copula. If one is interested in estimating the dependence parameter, this procedure reduces the computational time, while providing a similar accuracy as the conditional maximum likelihood estimation method for all of the model parameters.

Secondly, a more general specification for an integer-valued process for seasonality is also considered. More specifically, a univariate integer-valued autoregressive process for seasonality with period d ($\mathrm{SINAR}(1)_d$) is introduced, which allows the intra-seasonal dependence of the innovations to be described by a copula. Furthermore, it can be shown that such a univariate process can also be written as a multivariate specification, for which the BINAR(1) is a special case by taking the seasonal period $d = 2$. The multivariate specification allows the use of an estimation method based on least-squares, which can be generalized to account for the residual dependence. This method is then compared with a two-step likelihood-based estimation method, where emphasis is now put on the accuracy of the parameters of the $\mathrm{SINAR}(1)_d$ model itself, rather than the dependence parameter only.

Finally, the estimation methods of the above processes are illustrated using Monte Carlo simulation. Furthermore, the processes studied in this thesis are illustrated using empirical applications in the context of loan defaults (for the BINAR(1) process) and crime data (for the $\text{SINAR}(1)_d$ process).

Bibliography

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

Athreya, K. B. and Ney, P. E. (1972). *Branching Processes.* Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer-Verlag.

Barczy, M., Ispány, M., Pap, G., Scotto, M., and Silva, M. E. (2010). Innovational outliers in INAR(1) models. *Communications in Statistics - Theory and Methods*, 39(18):3343–3362.

Barreto-Souza, W. (2017). Mixed Poisson INAR(1) processes. *Statistical Papers*, 60(6):2119–2139.

Bourguignon, M., Vasconcellos, K. L., Reisen, V. A., and Ispány, M. (2016). A Poisson INAR(1) process with a seasonal structure. *Journal of Statistical Computation and Simulation*, 86(2):373–387.

Bourguignon, M. and Weiß, C. H. (2017). An INAR(1) process for modeling count time series with equidispersion, underdispersion and overdispersion. *TEST*, 26(4):847–868.

Brännäs, K. (1995). Explanatory variables in the AR(1) count data model. *Umeå Economic Studies 381.*

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

Buteikis, A. and Leipus, R. (2019). A copula-based bivariate integer-valued autoregressive process with application. *Modern Stochastics: Theory and Applications*, 6(2):227–249.

Buteikis, A. and Leipus, R. (2020). An integer-valued autoregressive process for seasonality. *Journal of Statistical Computation and Simulation*, 90(3):391–411.

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

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

Davis, R., Holan, S., Lund, R., and Ravishanker, N. (2016). *Handbook of Discrete-Valued Time Series*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press.

Ding, X. and Wang, D. (2016). Empirical likelihood inference for INAR(1) model with explanatory variables. *Journal of the Korean Statistical Society*, 45(4):623–632.

Efron, B. (1975). Biased versus unbiased estimation. *Advances in Mathematics*, 16(3):259 – 277.

Enciso-Mora, V., Neal, P., and Rao, T. S. (2009). Integer valued AR processes with explanatory variables. *Sankhyā: The Indian Journal of Statistics, Series B*, 71(2):248–263.

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

Franke, J. and Seligmann, T. (1993). Conditional maximum likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In Subba Rao, T., editor, *Developments in Time Series Analysis*, pages 310–330. Chapman & Hall: London.

Freeland, R. K. and McCabe, B. (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics & Probability Letters*, 73(2):147 – 153.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *Astin Bulletin*, 37(2):475–515.

Gladyshev, E. G. (1961). Periodically correlated random sequences. *Sov. Math., Dokl.*, 2:385–388.

Hurd, H. and Miamee, A. (2007). *Periodically Correlated Random Sequences: Spectral Theory and Practice.* Wiley Series in Probability and Statistics. Wiley.

Ispány, M., Pap, G., and van Zuijlen, M. C. A. (2003). Asymptotic behaviour of estimators of the parameters of nearly unstable INAR(1) models. In Haitovsky, Y., Ritov, Y., and Lerche, H. R., editors, *Foundations of Statistical Inference*, pages 194–206, Heidelberg. Physica-Verlag HD.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419.

Joe, H. (2015). *Dependence Modeling with Copulas.* Chapman & Hall/CRC Monographs on Statistics and Applied probability 134.

Joe, H. (2019). Likelihood inference for generalized integer autoregressive time series models. *Econometrics*, 7:43.

Kang, Y., Wang, D., Yang, K., and Zhang, Y. (2020). A new thinning-based INAR(1) process for underdispersed or overdispersed counts. *Journal of the Korean Statistical Society*, 49(2):324–349.

Karlis, D. and Pedeli, X. (2013). Flexible bivariate INAR(1) processes using copulas. *Communications in Statistics - Theory and Methods*, 42:723–740.

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

Kim, H.-Y. and Park, Y. (2006). A non-stationary integer-valued autoregressive model. *Statistical Papers*, 49(3):485–502.

Latour, A. (1997). The multivariate GINAR(p) process. *Advances in Applied Probability*, 29(1):228–248.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, 19(4):439–455.

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis.* Springer, Berlin Heidelberg.

Manstavičius, M. and Leipus, R. (2017). Bounds for the Clayton copula. *Nonlinear Analysis: Modelling and Control*, 22:248–260.

McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, 18(3):679–705.

Monteiro, M., Scotto, M. G., and Pereira, I. (2010). Integer-valued autoregressive processes with periodic structure. *Journal of Statistical Planning and Inference*, 140(6):1529–1541.

Nadarajah, S., Afuecheta, E., and Chan, S. (2018). A compendium of copulas. *Statistica*, 77:279–328.

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

Nikoloulopoulos, A. K. and Karlis, D. (2009). Modeling multivariate count data using copulas. *Communications in Statistics - Simulation and Computation*, 39(1):172–187.

Panagiotelis, A., Czado, C., and Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, 107(499):1063–1072.

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

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

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

Ristić, M. M., Bakouch, H. S., and Nastić, A. S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, 139(7):2218 – 2226.

Scotto, M., Weiß, C., and Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, 15.

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

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231.

Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Probab.*, 7(5):893–899.

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

Weiß, C. (2008). Thinning operations for modeling time series of counts - a survey. *AStA Advances in Statistical Analysis*, 92:319–341.

Weiß, C. (2018). *An Introduction to Discrete-Valued Time Series.* John Wiley & Sons, Ltd.

Appendix A

Proof of thinning operator properties

Below we present the proof of various thinning operator properties.

A.1 Proof of Theorem 2.1.1

Below we present the proofs of Theorem 2.1.1 for the thinning operator properties for the univariate setting.

2.1.1 (a) Let $\widetilde{B}_i \sim Bern(\alpha_1)$, $B_i \sim Bern(\alpha_2)$ and $\overline{B}_i \sim Bern(\alpha_1\alpha_2)$, $i = 1,2,...$. 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

$$\alpha_1 \circ (\alpha_2 \circ X) = \alpha_1 \circ \sum_{i=1}^{X} B_i = \sum_{j=1}^{\sum_{i=1}^{X} B_i} \widetilde{B}_j, \tag{A.1.1}$$

$$(\alpha_1\alpha_2) \circ X = \sum_{i=1}^{X} \overline{B}_i. \tag{A.1.2}$$

The characteristic function of equation (A.1.2) is:

$$\begin{aligned} \overline{\varphi}(u) &= \mathbb{E}\mathrm{e}^{\mathrm{i}u\sum_{i=1}^{X}\overline{B}_i} = \sum_{k=0}^{\infty} \mathbb{E}\mathrm{e}^{\mathrm{i}u\sum_{i=1}^{k}\overline{B}_i}\mathbb{P}(X=k) = \sum_{k=0}^{\infty} \overline{\varphi}^k_{\alpha_1\alpha_2}(u)\mathbb{P}(X=k) \\ &= \sum_{k=0}^{\infty} (\mathrm{e}^{\mathrm{i}u}\alpha_1\alpha_2 + (1-\alpha_1\alpha_2))^k \mathbb{P}(X=k), \end{aligned} \tag{A.1.3}$$

because

$$\overline{\varphi}_{\alpha_1\alpha_2}(u) = \mathbb{E}e^{iu\overline{B}} = \mathbb{E}(e^{iu\overline{B}}|\overline{B}=1)\mathbb{P}(\overline{B}=1) + \mathbb{E}(e^{iu\overline{B}}|\overline{B}=0)\mathbb{P}(\overline{B}=0)$$
$$= e^{iu}\alpha_1\alpha_2 + (1-\alpha_1\alpha_2).$$

The characteristic function of equation (A.1.1) is:

$$\widetilde{\varphi}(u) = \mathbb{E}e^{iu\sum_{j=1}^{\sum_{i=1}^{X} B_i}\widetilde{B}_j} = \sum_{k=0}^{\infty}\mathbb{E}e^{iu\sum_{j=1}^{\sum_{i=1}^{k} B_i}\widetilde{B}_j}\mathbb{P}(X=k). \tag{A.1.4}$$

We see that

$$\mathbb{E}e^{iu\sum_{j=1}^{\sum_{i=1}^{k} B_i}\widetilde{B}_j} = \sum_{l=0}^{k}\mathbb{E}e^{iu\sum_{j=1}^{l}\widetilde{B}_j}\mathbb{P}\left(\sum_{i=1}^{k}B_i = l\right) = \sum_{l=0}^{k}\varphi_{\alpha_1}^{l}(u)\mathbb{P}\left(\sum_{i=1}^{k}B_i = l\right), \tag{A.1.5}$$

where

$$\varphi_{\alpha_1}(u) = \mathbb{E}e^{iu\widetilde{B}} = \mathbb{E}(e^{iu\widetilde{B}}|\widetilde{B}=1)\mathbb{P}(\widetilde{B}=1) + \mathbb{E}(e^{iu\widetilde{B}}|\widetilde{B}=0)\mathbb{P}(\widetilde{B}=0)$$
$$= e^{iu}\alpha_1 + (1-\alpha_1),$$

and

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

From (A.1.4) and (A.1.5) we have that:

$$\widetilde{\varphi}(u) = \mathbb{E}e^{iu\sum_{j=1}^{\sum_{i=1}^{X} B_i}\widetilde{B}_j}$$
$$= \sum_{k=0}^{\infty}\sum_{l=0}^{k}(e^{iu}\alpha_1 + (1-\alpha_1))^l\binom{k}{l}\alpha_2^l(1-\alpha_2)^{k-l}\mathbb{P}(X=k) \tag{A.1.6}$$

Because $(a+b)^k = \sum_{l=0}^{k}\binom{k}{l}a^l b^{k-l}$ where $a, b \in \mathbb{C}$, 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 (A.1.6) becomes:

$$\begin{aligned}\widetilde{\varphi}(u) &= \sum_{k=0}^{\infty}(\mathrm{e}^{\mathrm{i}u}\alpha_1\alpha_2+(1-\alpha_1)\alpha_2+(1-\alpha_2))^k\mathbb{P}(X=k)\\ &= \sum_{k=0}^{\infty}(\mathrm{e}^{\mathrm{i}u}\alpha_1\alpha_2+(1-\alpha_1\alpha_2))^k\mathbb{P}(X=k). \qquad \text{(A.1.7)}\end{aligned}$$

Comparing equations (A.1.3) and (A.1.7) we see that they are equal, so we have that $\alpha_1 \circ (\alpha_2 \circ X) \overset{d}{=} (\alpha_1\alpha_2) \circ X$.

Alternatively, as mentioned by the reviewer, if the non-negative integer-valued r.v. X has the probability generating function (pgf) G, then the pgf of $\alpha \circ X$ can be derived as the composite function $G(1+\alpha(z-1))$, where $1+\alpha(z-1)$ is the pgf of the Bernoulli distribution with mean α, see Weiß (2018, p. 19). This is a result for branching processes, see Athreya and Ney (1972, p. 263).

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

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

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

$$\mathbb{E}\mathrm{e}^{\mathrm{i}u\sum_{i=1}^{k}B_i} = (\mathbb{E}\mathrm{e}^{\mathrm{i}uB_i})^k = (\mathrm{e}^{\mathrm{i}u}\alpha+(1-\alpha))^k.$$

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

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

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

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

Since B_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:

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

Thus, the equality in 2.1.1 (b) holds.

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

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

2.1.1 (d) We know that $\mathbb{V}\mathrm{ar}(\alpha \circ X) = \mathbb{E}(\alpha \circ X)^2 - (\mathbb{E}(\alpha \circ X))^2$. Using part 2.1.1 (c), we have that the second term can be expressed as

$$(\mathbb{E}(\alpha \circ X))^2 = (\alpha\mathbb{E}(X))^2 = \alpha^2(\mathbb{E}(X))^2, \tag{A.1.8}$$

and the first term as

$$\begin{aligned}\mathbb{E}(\alpha \circ X)^2 &= \mathbb{E}\left(\sum_{i=1}^{X} B_i\right)^2 = \mathbb{E}\left(\sum_{i,j=1}^{X} B_i B_j\right) = \mathbb{E}\left(\sum_{i\neq j} B_i B_j + \sum_{i=1}^{X} B_i^2\right) \\ &= \sum_{k=0}^{\infty} \mathbb{E}\left(\sum_{i\neq j} B_i B_j + \sum_{i=1}^{k} B_i^2\right)\mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty}\left(\sum_{i\neq j} \alpha^2 + \sum_{i=1}^{k} \alpha\right)\mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty}((k^2 - k)\alpha^2 + k\alpha)\mathbb{P}(X = k) \\ &= \alpha^2\sum_{k=0}^{\infty} k^2\mathbb{P}(X = k) + \alpha(1-\alpha)\sum_{k=0}^{\infty} k\mathbb{P}(X = k) \\ &= \alpha^2\mathbb{E}(X^2) + \alpha(1-\alpha)\mathbb{E}(X).\end{aligned} \tag{A.1.9}$$

From equations (A.1.8) and (A.1.9) we get

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

2.1.1 (e) We have that

$$
\begin{aligned}
\mathbb{E}(X_2^p(\alpha \circ X_1)) &= \mathbb{E}\left(\mathbb{E}((\alpha \circ X_1)X_2^p | X_1, X_2)\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{X_1} B_i X_2^p \middle| X_1, X_2\right)\right) = \mathbb{E}\left(\sum_{i=1}^{X_1} \mathbb{E}\left(B_i X_2^p \middle| X_1, X_2\right)\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{X_1} \mathbb{E}\left(B_i | X_1, X_2\right) X_2^p\right) = \mathbb{E}\left(X_2^p \sum_{i=1}^{X_1} \alpha\right) \\
&= \alpha \mathbb{E}(X_1 X_2^p).
\end{aligned}
$$

2.1.1 (f) Using part 2.1.1 (e) with $X_1 = X_2 = X$, we have that the equality $\mathbb{E}\left[X^p(\alpha \circ X)\right] = \alpha \mathbb{E}(X^{p+1})$ holds true.

2.1.1 (g) Using parts 2.1.1 (c) and 2.1.1 (e), we get that

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

2.1.1 (h) Similarly to part 2.1.1 (e), we have that

$$
\begin{aligned}
\mathbb{E}((\alpha_1 \circ X_1)(\alpha_2 \circ X_2)) &= \mathbb{E}\left(\mathbb{E}((\alpha \circ X_1)(\alpha_2 \circ X_2) | X_1, X_2)\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{X_1} \sum_{j=1}^{X_2} B_i B_j | X_1, X_2\right)\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{X_1} \sum_{j=1}^{X_2} \mathbb{E}\left(B_i B_j | X_1, X_2\right)\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{X_1} \sum_{j=1}^{X_2} \mathbb{E}\left(B_i | X_1, X_2\right) \mathbb{E}\left(B_j | X_1, X_2\right)\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{X_1} \sum_{j=1}^{X_2} \alpha_1 \alpha_2\right) \\
&= \alpha_1 \alpha_2 \mathbb{E}(X_1 X_2).
\end{aligned}
$$

2.1.1 (i) Similarly to eq. (A.1.9) and part 2.1.1 (f), we have that

$$\begin{aligned}
\mathbb{E}\left[X^p(\alpha\circ X)^2\right] &= \mathbb{E}X^p\left(\sum_{i=1}^{X} B_i\right)^2 = \mathbb{E}\left[\left(X^p\sum_{i,j=1}^{X} B_iB_j\right)\right]\\
&= \mathbb{E}\left(X^p\left[\sum_{i\neq j} B_iB_j + \sum_{i=1}^{X} B_i^2\right]\right)\\
&= \sum_{k=0}^{\infty}\mathbb{E}\left(k^p\left[\sum_{i\neq j} B_iB_j + \sum_{i=1}^{k} B_i^2\right]\right)\mathbb{P}(X=k)\\
&= \sum_{k=0}^{\infty}\left(k^p\left[\sum_{i\neq j} \alpha^2 + \sum_{i=1}^{k} \alpha\right]\right)\mathbb{P}(X=k)\\
&= \sum_{k=0}^{\infty} k^p((k^2-k)\alpha^2+k\alpha)\mathbb{P}(X=k)\\
&= \alpha^2\sum_{k=0}^{\infty} k^{p+2}\mathbb{P}(X=k)+\alpha(1-\alpha)\sum_{k=0}^{\infty} k^{p+1}\mathbb{P}(X=k)\\
&= \alpha^2\mathbb{E}(X^{p+2})+\alpha(1-\alpha)\mathbb{E}(X^{p+1}).
\end{aligned}$$

A.2 Proof of Theorem 2.2.1

Below we present the proofs of Theorem 2.2.1 for the thinning operator properties for the multivariate setting.

2.2.1 (a) The m-th element of $\mathbf{A}_j\circ\mathbf{X}_j$ is $\sum_{s=1}^{k}\alpha_{m,s,j}\circ X_{s,j}$, then by applying property 2.1.1 (c) to each element, it holds that

$$\mathbb{E}[\mathbf{A}_j\circ\mathbf{X}_j] = \begin{bmatrix}\sum_{s=1}^{k}\alpha_{1,s,j}\mathbb{E}[X_{s,j}] & \dots & \sum_{s=1}^{k}\alpha_{k,s,j}\mathbb{E}[X_{s,j}]\end{bmatrix}^\top = \mathbf{A}_j\mathbb{E}[\mathbf{X}_j].$$

2.2.1 (b) The (m_1,m_2)-th element of $[\mathbf{A}_i\circ\mathbf{X}_i][\mathbf{A}_j\circ\mathbf{X}_j]^\top$ is

$$\sum_{s_1=1}^{k}\sum_{s_2=1}^{k}(\alpha_{m_1,s_1,i}\circ X_{s_1,i})(\alpha_{m_2,s_2,j}\circ X_{s_2,j}).$$

From property 2.1.1 (h) and eq. (A.1.9) the expected value of the

(m_1, m_2)-th element is

$$\sum_{s_1=1}^{k}\sum_{s_2=1}^{k}\mathbb{E}(\alpha_{m_1,s_1,i}\circ X_{s_1,i})(\alpha_{m_2,s_2,j}\circ X_{s_2,j})$$
$$= \sum_{s_1=1}^{k}\sum_{s_2=1}^{k}\alpha_{m_1,s_1,i}\alpha_{m_2,s_2,j}\mathbb{E}[X_{s_1,i}X_{s_2,j}]$$
$$+ \mathbb{1}_{\{(m_1=m_2)\cap(i=j)\}}\sum_{s_1=1}^{k}\alpha_{m_1,s_1,i}(1-\alpha_{m_1,s_1,i})\mathbb{E}[X_{s_1,i}].$$

It then holds that

$$\mathbb{E}[\mathbf{A}_i\circ\mathbf{X}_i][\mathbf{A}_j\circ\mathbf{X}_j]^\top = \mathbf{A}_i\mathbb{E}[\mathbf{X}_i\mathbf{X}_j^\top]\mathbf{A}_j^\top + \mathbb{1}_{\{i=j\}}\text{diag}(\mathbf{B}_i\mathbb{E}[\mathbf{X}_i]),$$

where $\mathbf{B}_i = (\alpha_{m_1,m_2,i}(1-\alpha_{m_1,m_2,i}))_{m_1,m_2=1,\ldots,k}$.

2.2.1 (c) The (m_1, m_2)-th element of $[\mathbf{A}_i\circ\mathbf{X}_i][\mathbf{A}_j\mathbf{X}_j]^\top$ is $\sum_{s_1=1}^{k}\sum_{s_2=1}^{k}(\alpha_{m_1,s_1,i}\circ X_{s_1,i})(\alpha_{m_2,s_2,j}X_{s_2,j})$ similarly to 2.1.1 (e) and 2.1.1 (f), the expected value of the (m_1, m_2)-th element is

$$\sum_{s_1=1}^{k}\sum_{s_2=1}^{k}\alpha_{m_2,s_2,j}\mathbb{E}[(\alpha_{m_1,s_1,i}\circ X_{s_1,i})X_{s_2,j}] = \sum_{s_1=1}^{k}\sum_{s_2=1}^{k}\alpha_{m_1,s_1,i}\alpha_{m_2,s_2,j}\mathbb{E}[X_{s_1,i}X_{s_2,j}].$$

It then holds that $\mathbb{E}[\mathbf{A}_i\circ\mathbf{X}_i][\mathbf{A}_j\mathbf{X}_j]^\top = \mathbf{A}_i\mathbb{E}[\mathbf{X}_i\mathbf{X}_j^\top]\mathbf{A}_j^\top$.

2.2.1 (d) As the (m_1, m_2)-th element of $[\mathbf{A}_i\mathbf{X}_i][\mathbf{A}_j\circ\mathbf{X}_j]^\top$ is $\sum_{s_1=1}^{k}\sum_{s_2=1}^{k}(\alpha_{m_1,s_1,i}X_{s_1,i})(\alpha_{m_2,s_2,j}\circ X_{s_2,j})$, the proof is analogous to 2.2.1 (c).

Appendix B

Proof of BINAR(1) properties

B.1 Proof of Theorem 4.1.1

Below we present the proofs of Theorem 4.1.1 for the properties of the BINAR(1) process.

We would also like to mention that the expected value in 4.1.1 (a) and the covariances in 4.1.1 (d) and 4.1.1 (f) can be calculated recursively, as mentioned by the reviewer.

4.1.1 (a) We have

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

Here, first equality is from the definition of BINAR(1) model from equation (4.1.1). We get the second, third and fourth equalities by using the definition of BINAR(1) model expressed in terms of arrival processes (4.1.3) and properties 2.1.1 (a) and 2.1.1 (b). The fifth equality is from property 2.1.1 (c), as well as Fubini's theorem, since the infinite sum in eq. (4.1.3) converges a.s. with $\alpha_j \in [0,1)$. The last equality is from the definition of an infinite geometric series.

4.1.1 (b) From the 2.1.1 (e) property of the binomial thinning operator and Def. 4.1.1 it follows that

$$\begin{aligned}
\mathbb{E}(Y_{j,t}|Y_{j,t-1}) &= \mathbb{E}(\alpha_j \circ Y_{j,t-1} + \varepsilon_{j,t}|Y_{j,t-1}) \\
&= \mathbb{E}(\alpha_j \circ Y_{j,t-1}|Y_{j,t-1}) + \mathbb{E}(\varepsilon_{j,t}|Y_{j,t-1}) \\
&= \alpha_j \mathbb{E}(Y_{j,t-1}|Y_{j,t-1}) + \mathbb{E}(\varepsilon_{j,t}) \\
&= \alpha_j Y_{j,t-1} + \mu_{\varepsilon,j}.
\end{aligned}$$

4.1.1 (c) We have

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

Here, the first equality is from equation (4.1.3). The second equality is from Fubini's theorem and the fact that $\varepsilon_{j,t-k}$ are i.i.d. The third equality is from 2.1.1 (d).

4.1.1 (d) We have that

$$\begin{aligned}
\mathbb{C}\text{ov}(Y_{i,t},\varepsilon_{j,t}) &= \mathbb{E}(Y_{i,t}\varepsilon_{j,t}) - \mathbb{E}(Y_{i,t})\mathbb{E}(\varepsilon_{j,t}) \\
&= \mathbb{E}(Y_{i,t}\varepsilon_{j,t}) - \mu_{Y_i}\mu_{\varepsilon,j} \\
&= \mathbb{E}\left(\left(\sum_{k=0}^{\infty}\alpha_i^k \circ \varepsilon_{i,t-k}\right)\varepsilon_{j,t}\right) - \mu_{Y_i}\mu_{\varepsilon,j} \\
&= \sum_{k=0}^{\infty}\alpha_i^k \mathbb{E}(\varepsilon_{i,t-k}\varepsilon_{j,t}) - \mu_{Y_i}\mu_{\varepsilon,j} \\
&= \mathbb{E}(\varepsilon_{i,t}\varepsilon_{j,t}) + \sum_{k=1}^{\infty}\alpha_i^k \mathbb{E}(\varepsilon_{i,t-k}\varepsilon_{j,t}) - \mu_{Y_i}\mu_{\varepsilon,j} \\
&= \mathbb{E}(\varepsilon_{i,t}\varepsilon_{j,t}) + \sum_{k=1}^{\infty}\alpha_i^k \mu_{\varepsilon,i}\mu_{\varepsilon,j} - \frac{\mu_{\varepsilon,i}\mu_{\varepsilon,j}}{1-\alpha_i} \\
&= \mathbb{C}\text{ov}(\varepsilon_{i,t},\varepsilon_{j,t}).
\end{aligned}$$

Here we use Fubini's theorem, eq. (4.1.3), the fact that $\varepsilon_{j,t-k}$ are i.i.d. in t and 4.1.1 (a).

4.1.1 (e) We have that

$$\begin{aligned}
\mathbb{C}\text{ov}(Y_{j,t},Y_{j,t+h}) &= \mathbb{C}\text{ov}(Y_{j,t},\alpha_j^h \circ Y_{j,t} + \sum_{k=0}^{h-1}\alpha_j^k \circ \varepsilon_{j,t+h-k}) \\
&= \mathbb{C}\text{ov}(Y_{j,t},\alpha_j^h \circ Y_{j,t}) + \mathbb{C}\text{ov}(Y_{j,t},\sum_{k=0}^{h-1}\alpha_j^k \circ \varepsilon_{j,t+h-k}) \\
&= \alpha_j^h \mathbb{C}\text{ov}(Y_{j,t},Y_{j,t}) = \alpha_j^h \sigma_{Y_j}^2.
\end{aligned}$$

Here, using the fact that $\varepsilon_{j,t}$ are i.i.d in t and $t+h-k>t$ for $k<h$, we have that $\mathbb{C}\text{ov}(Y_{j,t},\sum_{k=0}^{h-1}\alpha_j^k \circ \varepsilon_{j,t+h-k}) = 0$. We have also used Theorem 2.1.1 (g) to get the next-to-last equality.

4.1.1 (f) Using covariance and variance expressions from 4.1.1 (c) and 4.1.1 (e) we have that

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

4.1.1 (g) Using equations (4.1.2), (4.1.3) and Theorem 2.1.1 we have that

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

because:

$$\begin{aligned}
\sum_{l=0}^{h-1} \alpha_j^l \mathbb{E}\left(Y_{i,t}\varepsilon_{j,t+h-l}\right) &= \sum_{l=0}^{h-1} \alpha_j^l \mathbb{E}(Y_{i,t})\mathbb{E}(\varepsilon_{j,t+h-l}) \\
&= \mu_{Y_i}\mu_{\varepsilon,j} \left(\sum_{l=0}^{h-1} \alpha_j^l \right) \\
&= \mu_{Y_i}\mu_{\varepsilon,j} \frac{1-\alpha_j^h}{1-\alpha_j} \\
&= (1-\alpha_j^h)\mu_{Y_i}\mu_{Y_j},
\end{aligned}$$

and

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

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.

4.1.1 (h) From the definition of the correlation as well as using properties 4.1.1 (c) and 4.1.1 (g) we have that

$$\begin{aligned}\mathbb{C}\text{orr}(Y_{i,t+h}, Y_{j,t}) &= \frac{\mathbb{C}\text{ov}(Y_{i,t+h}, Y_{j,t})}{\sqrt{\mathbb{V}\text{ar}(Y_{i,t+h})\mathbb{V}\text{ar}(Y_{j,t})}} \\ &= \frac{\alpha_i^h \sqrt{(1-\alpha_i^2)(1-\alpha_j^2)}}{(1-\alpha_1\alpha_2)\sqrt{(\sigma_i^2+\alpha_i\mu_{\varepsilon,i})(\sigma_j^2+\alpha_j\mu_{\varepsilon,j})}} \mathbb{C}\text{ov}(\varepsilon_{i,t}, \varepsilon_{j,t})\end{aligned}$$

holds for $h \geq 0$.

B.2 Proof of equation (4.2.10)

We provide the derivation of equation (4.2.10). Let $F_{j,k} = F_j(k; \mu_{\varepsilon,j})$, $\overline{F}_{j,k} = 1 - F_{j,k}$, $j = 1,2$. Using eq. (3.0.5), eq. (4.2.6) and eq. (4.2.7) we have that the following holds:

$$\begin{aligned}\mathbb{E}\left(\varepsilon_{1,t}\varepsilon_{2,t}\right) &= \sum_{k=1}^{\infty}\sum_{s=1}^{\infty} ks\left(F_{1,k}F_{2,s}\left(1+\theta\overline{F}_{1,k}\overline{F}_{2,s}\right) - F_{1,k}F_{2,s-1}\left(1+\theta\overline{F}_{1,k}\overline{F}_{2,s-1}\right)\right. \\ &\quad \left. - F_{1,k-1}F_{2,s}\left(1+\theta\overline{F}_{1,k-1}\overline{F}_{2,s}\right) + F_{1,k-1}F_{2,s-1}\left(1+\theta\overline{F}_{1,k-1}\overline{F}_{2,s-1}\right)\right) \\ &= \mathbb{E}\varepsilon_{1,t}\mathbb{E}\varepsilon_{2,s} + \theta\sum_{k=1}^{\infty}\sum_{s=1}^{\infty} ksF_{1,k}F_{2,s}\overline{F}_{1,k}\overline{F}_{2,s} \\ &\quad - \theta\sum_{k=1}^{\infty}\sum_{s=1}^{\infty} ksF_{1,k}F_{2,s-1}\overline{F}_{1,k}\overline{F}_{2,s-1} - \theta\sum_{k=1}^{\infty}\sum_{s=1}^{\infty} ksF_{1,k-1}F_{2,s}\overline{F}_{1,k-1}\overline{F}_{2,s} \\ &\quad + \theta\sum_{k=1}^{\infty}\sum_{s=1}^{\infty} ksF_{1,k-1}F_{2,s-1}\overline{F}_{1,k-1}\overline{F}_{2,s-1} \\ &= \mathbb{E}\varepsilon_{1,t}\mathbb{E}\varepsilon_{2,s} + \theta\sum_{k=1}^{\infty} k\left(F_{1,k}\overline{F}_{1,k} - F_{1,k-1}\overline{F}_{1,k-1}\right) \\ &\quad \times \sum_{s=1}^{\infty} s\left(F_{2,s}\overline{F}_{2,s} - F_{s,2-1}\overline{F}_{2,s-1}\right)\end{aligned}$$

Substituting the above expression in eq. (4.2.6) we get:

$$\gamma(\mu_{\varepsilon,1}, \mu_{\varepsilon,2}; \theta) = \theta\sum_{k=1}^{\infty} k\left(F_{1,k}\overline{F}_{1,k} - F_{1,k-1}\overline{F}_{1,k-1}\right)\sum_{s=1}^{\infty} s\left(F_{2,s}\overline{F}_{2,s} - F_{s,2-1}\overline{F}_{2,s-1}\right)$$

Inserting the new expression of $\gamma(\mu_{\varepsilon,1}, \mu_{\varepsilon,2}; \theta)$ in equation (4.2.5) and taking the derivative with respect to θ and equating it to zero gives us:

$$
\begin{aligned}
(N-1)\hat{\theta}_{\mathrm{FGM}}^{\mathrm{CLS}} \sum_{k=1}^{\infty} k \left(F_{1,k}\overline{F}_{1,k} - F_{1,k-1}\overline{F}_{1,k-1}\right) \sum_{s=1}^{\infty} s \left(F_{2,s}\overline{F}_{2,s} - F_{s,2-1}\overline{F}_{2,s-1}\right) \\
= \sum_{t=2}^{N} \left((Y_{1,t} - \hat{\alpha}_1^{\mathrm{CLS}} Y_{1,t-1} - \hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}})(Y_{2,t} - \hat{\alpha}_2^{\mathrm{CLS}} Y_{2,t-1} - \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}})\right)
\end{aligned}
$$

Rearranging the equality we get:

$$
\hat{\theta}_{\mathrm{FGM}}^{\mathrm{CLS}} = \frac{\sum_{t=2}^{N} (Y_{1,t} - \hat{\alpha}_1^{\mathrm{CLS}} Y_{1,t-1} - \hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}})(Y_{2,t} - \hat{\alpha}_2^{\mathrm{CLS}} Y_{2,t-1} - \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}})}{(N-1)\sum_{k=1}^{\infty} k \left(F_{1,k}\overline{F}_{1,k} - F_{1,k-1}\overline{F}_{1,k-1}\right) \sum_{s=1}^{\infty} s \left(F_{2,s}\overline{F}_{2,s} - F_{2,s-1}\overline{F}_{2,s-1}\right)}
$$

Using the approximation of the covariance function in eq. (4.2.8) completes the proof.

Appendix C

Inference on the estimate bias from the BINAR(1) Monte Carlo simulation

Let our simulated data in the Monte-Carlo simulation results be $X_{j,1}^{(i)},...,X_{j,N}^{(i)}$ for simulated sample $i = 1,...,M$ and $j = 1,2$. Let $\eta \in \{\alpha_1, \alpha_2, \mu_{\varepsilon,1}, \mu_{\varepsilon,2}, \theta, \sigma_2^2\}$ and let $\widehat{\eta}^{(i)}$ be either a CLS, CML or Two-step estimate of the true parameter value η for the simulated sample i.

The mean squared error and the bias are calculated as follows:

$$\text{MSE}(\widehat{\eta}) = \frac{1}{M}\sum_{i=1}^{M}(\widehat{\eta}^{(i)} - \eta)^2,$$
$$\text{Bias}(\widehat{\eta}) = \frac{1}{M}\sum_{i=1}^{M}(\widehat{\eta}^{(i)} - \eta).$$

Calculating the per-sample bias for each simulated sample i would also allow us to calculate the sample variance of biases $\text{Bias}(\widehat{\eta}^{(i)}) = \widehat{\eta}^{(i)} - \eta$, $i = 1,...,M$:

$$\widehat{\mathbb{V}\text{ar}}\left(\text{Bias}(\widehat{\eta})\right) = \frac{1}{M-1}\sum_{i=1}^{M}\left[\text{Bias}(\widehat{\eta}^{(i)}) - \text{Bias}(\widehat{\eta})\right]^2,$$

which we can use to calculate the standard error of the bias.

The standard errors of the parameter bias of the Monte Carlo simulation are presented in Table C.0.1. The columns labelled 'P-P' indicate the cases where both innovations have Poisson marginal distributions, while columns labelled 'P-NB' is for the cases where one innovation component follows Poisson and the other - a negative binomial distribution.

Table C.0.1: Standard errors of the bias of the estimated parameters from the Monte Carlo simulation.

Copula	Sample size	Parameter	True value	CLS		CML		Two-Step	
				P-P	P-NB	P-P	P-NB	P-P	P-NB
FGM	$N=50$	α_1	0.6	0.12396	0.12465	0.09252	0.09073	-	-
		α_2	0.4	0.13274	0.13029	0.12510	0.08541	-	-
		$\mu_{\varepsilon,1}$	1	0.33494	0.33631	0.25311	0.23225	-	-
		$\mu_{\varepsilon,2}$	2	0.48040	0.61210	0.44024	0.48916	-	-
		θ	-0.5	0.53139	0.54330	0.57707	0.53850	0.56899	0.53887
		σ_2^2	9	-	5.15368	-	3.88865	-	4.60221
	$N=500$	α_1	0.6	0.03813	0.03893	0.02706	0.02745	-	-
		α_2	0.4	0.04258	0.04392	0.03585	0.02306	-	-
		$\mu_{\varepsilon,1}$	1	0.10018	0.10076	0.07455	0.07367	-	-
		$\mu_{\varepsilon,2}$	2	0.15433	0.19676	0.13266	0.15377	-	-
		θ	-0.5	0.21631	0.25760	0.20666	0.20741	0.20657	0.20770
		σ_2^2	9	-	1.50841	-	1.34701	-	1.36119
Frank	$N=50$	α_1	0.6	0.12882	0.12975	0.09552	0.09420	-	-
		α_2	0.4	0.13158	0.13073	0.12448	0.08543	-	-
		$\mu_{\varepsilon,1}$	1	0.34266	0.34594	0.25719	0.23879	-	-
		$\mu_{\varepsilon,2}$	2	0.47982	0.61879	0.43939	0.48409	-	-
		θ	-1	1.34944	1.34314	1.43522	1.32589	1.40547	1.29538
		σ_2^2	9	-	4.98845	-	3.85654	-	4.62540
	$N=500$	α_1	0.6	0.03862	0.03951	0.02734	0.02727	-	-
		α_2	0.4	0.04212	0.04312	0.03591	0.02240	-	-
		$\mu_{\varepsilon,1}$	1	0.10091	0.10329	0.07409	0.07481	-	-
		$\mu_{\varepsilon,2}$	2	0.15490	0.19278	0.13351	0.15287	-	-
		θ	-1	0.46985	0.56268	0.44862	0.43540	0.44802	0.43627
		σ_2^2	9	-	1.31856	-	1.35359	-	1.36369
Clayton	$N=50$	α_1	0.6	0.12352	0.12684	0.08846	0.09360	-	-
		α_2	0.4	0.13123	0.12798	0.12361	0.07779	-	-
		$\mu_{\varepsilon,1}$	1	0.33505	0.33926	0.24609	0.24514	-	-
		$\mu_{\varepsilon,2}$	2	0.48252	0.62066	0.44194	0.48075	-	-
		θ	1	0.84763	0.88328	0.82176	0.79618	0.77890	0.75037
		σ_2^2	9	-	4.93912	-	3.91243	-	4.81812
	$N=500$	α_1	0.6	0.03782	0.03850	0.02641	0.02742	-	-
		α_2	0.4	0.04337	0.04410	0.03468	0.02176	-	-
		$\mu_{\varepsilon,1}$	1	0.09804	0.10033	0.07162	0.07180	-	-
		$\mu_{\varepsilon,2}$	2	0.15612	0.19969	0.13071	0.15185	-	-
		θ	1	0.33857	0.31212	0.23852	0.23316	0.23717	0.23476
		σ_2^2	9	-	1.57798	-	1.34163	-	1.37066

The results in Table C.0.1 are in line with the conclusions presented in Section 4.3 – for $\hat{\alpha}_j$, $\hat{\mu}_{\varepsilon,j}$, $j = 1,2$ and $\hat{\sigma}_2^2$ the standard error of the bias is smaller for CML compared to CLS. On the other hand, $\hat{\theta}$ has a similar standard error of the bias for CML and Two-step estimation method.

The kernel density estimate for the bias of the dependence parameter estimate, $\widehat{\theta}$, is presented in Figure C.0.1 for the Monte-Carlo simulation cases, where the sample size was 500. From Figure C.0.1 we see that the CML and Two-step estimates of the dependence parameter θ are similar to one another and have a lower standard error of the bias, compared to the CLS estimation method.

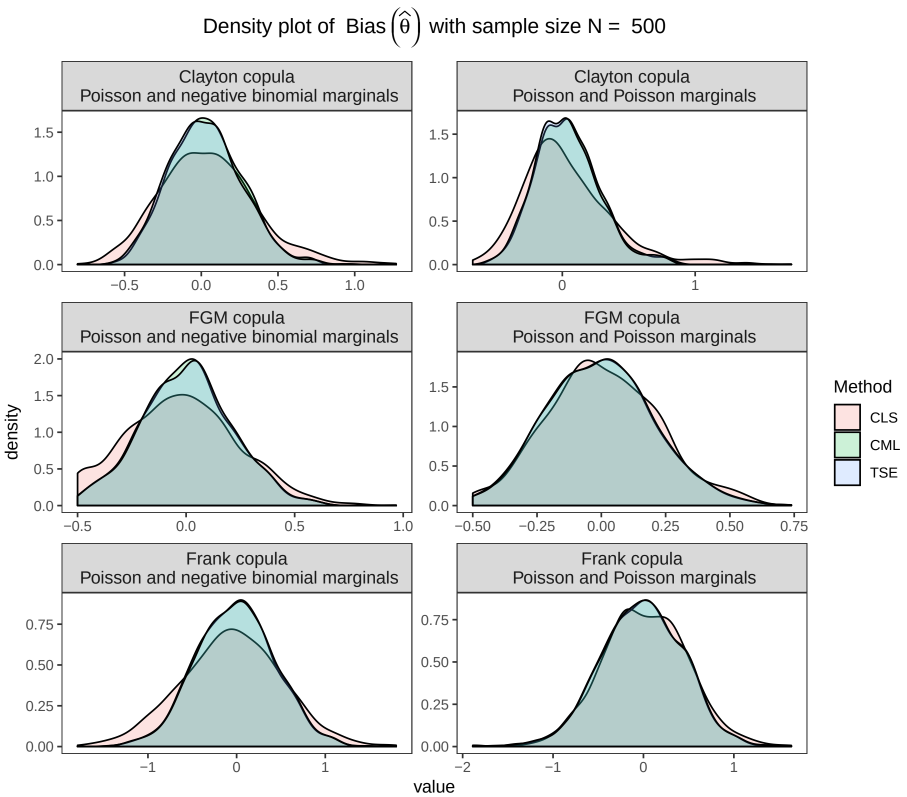


Figure C.0.1: Kernel density estimate of the bias of the dependence parameter estimates in the Monte Carlo simulation.

Appendix D

Proof of SINAR$(1)_d$ properties

D.1 Proof of Proposition 5.1.1

5.1.1 (a) Let $t = j + kd$, $k \in \mathbb{Z}$ and $j \in \{1,...,d\}$. Recursively applying eq. (5.1.1), we get that

$$\begin{aligned} Y_{j+kd} &= \phi_{j+kd} \circ Y_{j+(k-1)d} + \varepsilon_{j+kd} \overset{\mathrm{d}}{=} \left(\prod_{l=0}^{n} \phi_{j+(k-l)d} \right) \circ Y_{j+(k-n-1)d} \\ &+ \varepsilon_{j+kd} + \sum_{l=1}^{n} \left(\prod_{i=1}^{l} \phi_{j+(k-(i-1))d} \right) \circ \varepsilon_{j+(k-l)d} \\ &= \alpha_j^n \circ Y_{j+(k-n-1)d} + \sum_{l=0}^{n} \alpha_j^l \circ \varepsilon_{j+(k-l)d} \end{aligned} \tag{D.1.1}$$

for any $n \geq 1$, where we use the fact that $\phi_t = \alpha_j$, $t = j + kd$, $k \in \mathbb{Z}$. We can then continue the proof similarly to Bourguignon et al. (2016). For a fixed value $j \in \{1,\ldots,d\}$ and $k \in \mathbb{Z}$, we define the following random variable

$$X_{j,k,n} = \sum_{l=0}^{n} \alpha_j^l \circ \varepsilon_{j+(k-l)d},$$

where the counting sequences involved in $\alpha_j^l \circ \varepsilon_{j+(k-l)d}$ are the same for each fixed j and k, as $l \longrightarrow \infty$. Then, for all $0 < n < m$ and $k \in \mathbb{Z}$, we get

$$\begin{aligned}
\mathbb{E}\big(X_{j,k,m}-X_{j,k,n-1}\big) &= \mathbb{E}\Bigg(\sum_{l=n}^{m}\alpha_j^l\circ\varepsilon_{j+(k-l)d}\Bigg) = \sum_{l=n}^{m}\alpha_j^l\mu_{\varepsilon,j}\\
&= \frac{\alpha_j^n(1-\alpha_j^{m+1-n})}{1-\alpha_j}\mu_{\varepsilon,j}\\
&\le \frac{\alpha_j^n}{1-\alpha_j}\mu_{\varepsilon,j} \;\to\; 0 \text{ as } n\longrightarrow\infty,
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\big(X_{j,k,m}-X_{j,k,n-1}\big) &= \mathbb{V}\mathrm{ar}\Bigg(\sum_{l=n}^{m}\alpha_j^l\circ\varepsilon_{j+(k-l)d}\Bigg)\\
&= \sum_{l=n}^{m}\big(\alpha_j^{2l}\sigma_{\varepsilon,jj}+\alpha_j^l(1-\alpha_j^l)\mu_{\varepsilon,j}\big)\\
&\le \frac{\alpha_j^{2n}(1-\alpha_j^{2m+2-2n})}{1-\alpha_j^2}\sigma_{\varepsilon,jj}\\
&+ \frac{\alpha_j^n(1-\alpha_j^{m+1-n})}{1-\alpha_j}\mu_{\varepsilon,j}\\
&\le \frac{\alpha_j^{2n}}{1-\alpha_j^2}\sigma_{\varepsilon,jj}+\frac{\alpha_j^n}{1-\alpha_j}\mu_{\varepsilon,j} \;\to\; 0 \text{ as } n\to\infty,
\end{aligned}$$

which means that for each $j\in\{1,\dots,d\}$ and $k\in\mathbb{Z}$, $X_{j,k,n},\ n\ge 1$ forms a Cauchy sequence in the mean square sense and consequently, in probability. This implies that for a fixed $j\in\{1,\dots,d\}$ and any $k\in\mathbb{Z}$ it holds true that

$$X_{j,k,n}\xrightarrow{\text{P}}X_{j,k}\quad\text{as}\quad n\to\infty \tag{D.1.2}$$

with a random variable $X_{j,k}$ having the form of infinite sum on the r.h.s. of (5.1.2). Now, by (D.1.1), we can write $\{Y_t\}$ as d processes $\{Y_{j+kd},k\in\mathbb{Z}\},\ j=1,\dots,d$ of the form

$$Y_{j+kd} \;\stackrel{\text{d}}{=}\; \alpha_j^n\circ Y_{j+(k-n-1)d}+X_{j,k,n}.$$

Let the mean and variance of each of these stationary processes be defined as $\mu_{Y,j}<\infty$ and $\sigma_{Y,j}^2<\infty$, respectively. Since $\alpha_j\in[0,1)$, it holds that

$$\begin{aligned}
\mathbb{E}\left(\alpha_j^n \circ Y_{j+(k-n-1)d}\right) &= \alpha_j^n \mu_{Y,j} \ \rightarrow\ 0, \\
\mathbb{V}\mathrm{ar}\left(\alpha_j^n \circ Y_{j+(k-n-1)d}\right) &= \alpha_j^{2n}\sigma_{Y,j}^2 + \alpha_j^n(1-\alpha_j^n)\mu_{Y,j} \ \rightarrow\ 0 \text{ as } n \rightarrow \infty.
\end{aligned}$$

Therefore, for any k and $j \in \{1,...,d\}$, $\alpha_j^n \circ Y_{j+(k-n-1)d} \xrightarrow{\text{P}} 0$ as $n \rightarrow \infty$. This result, along with eq. (D.1.2) yields

$$Y_{j+kd} \overset{\text{d}}{=} X_{j,k} \text{ for all } k \in \mathbb{Z}.$$

This means that the unique stationary marginal distribution of process $\{Y_{j+kd} = \alpha_j \circ Y_{j+(k-1)d} + \varepsilon_{j+kd}, k \in \mathbb{Z}\}$ has the form in (5.1.2) (although it is not identical across $j = 1, \ldots, d$).

5.1.1 (b) By applying 5.1.1 (a) and the properties of the binomial thinning operator we have:

$$\mathbb{E}(Y_t) = \mathbb{E}\left(\sum_{l=0}^{\infty} \alpha_j^l \circ \varepsilon_{t-ld}\right) = \sum_{l=0}^{\infty} \alpha_j^l \mathbb{E}(\varepsilon_{t-ld}) = \frac{\mu_{\varepsilon,j}}{1-\alpha_j}.$$

5.1.1 (c) By applying 5.1.1 (a), the properties of the binomial thinning operator and using the fact that ε_t are only intra-seasonally-dependent, we have:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(Y_t) &= \mathbb{V}\mathrm{ar}\left(\sum_{l=0}^{\infty} \alpha_j^l \circ \varepsilon_{t-ld}\right) = \sum_{l=0}^{\infty} \mathbb{V}\mathrm{ar}\left(\alpha_j^l \circ \varepsilon_{t-ld}\right) \\
&= \sum_{l=0}^{\infty} \left(\alpha_j^{2l} \mathbb{V}\mathrm{ar}(\varepsilon_{t-ld}) + \alpha_j^l(1-\alpha_j^l)\mathbb{E}(\varepsilon_{t-ld})\right) \\
&= \sum_{l=0}^{\infty} \left(\alpha_j^{2l}\sigma_{\varepsilon,jj} + \alpha_j^l(1-\alpha_j^l)\mu_{\varepsilon,j}\right) = \frac{\sigma_{\varepsilon,jj} + \alpha_j \mu_{\varepsilon,j}}{1-\alpha_j^2}.
\end{aligned}$$

5.1.1 (d) If $s = j - i$ such that $i \in \{1, \ldots, d\}$ and $i \neq j$, then $t - s = i + kd$.

Because the process $\{Y_t, t \in j + \mathbb{Z}d\}$ is stationary, we have that

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(Y_{j+kd}, Y_{i+kd}) &= \mathbb{C}\mathrm{ov}(\alpha_j \circ Y_{j+(k-1)d} + \varepsilon_{j+kd}, \alpha_i \circ Y_{i+(k-1)d} + \varepsilon_{i+kd}) \\
&= \alpha_i \alpha_j \mathbb{C}\mathrm{ov}(Y_{j+(k-1)d}, Y_{i+(k-1)d}) + \sigma_{\varepsilon,ij} \\
&= \ldots \\
&= (\alpha_i \alpha_j)^{n+1} \mathbb{C}\mathrm{ov}(Y_{j+(k-(n+1))d}, Y_{i+(k-(n+1))d}) \\
&+ \sigma_{\varepsilon,ij} \sum_{l=0}^{n} (\alpha_i \alpha_j)^l, \text{ where } n \in \mathbb{N} \cup \{0\}.
\end{aligned}$$

Let $D := (\alpha_i \alpha_j)^{n+1} \mathbb{C}\mathrm{ov}(Y_{j+(k-(n+1))d}, Y_{i+(k-(n+1))d})$. Then the difference

$$\begin{aligned}
\left| D + \sum_{l=0}^{n} (\alpha_i \alpha_j)^l \sigma_{\varepsilon,ij} - \sigma_{\varepsilon,ij} \sum_{l=0}^{\infty} (\alpha_i \alpha_j)^l \right| &= \left| D - \sigma_{\varepsilon,ij} \sum_{l=n+1}^{\infty} (\alpha_i \alpha_j)^l \right| \\
&\leq (\alpha_i \alpha_j)^{n+1} \mathbb{E}(Y_{j+(k-(n+1))d} Y_{i+(k-(n+1))d}) + \sigma_{\varepsilon,ij} \frac{(\alpha_i \alpha_j)^{n+1}}{1 - \alpha_i \alpha_j} \\
&\leq (\alpha_i \alpha_j)^{n+1} \sqrt{\mathbb{E}(Y_j^2) \mathbb{E}(Y_i^2)} + \sigma_{\varepsilon,ij} \frac{(\alpha_i \alpha_j)^{n+1}}{1 - \alpha_i \alpha_j} \;\to\; 0 \text{ as } n \to \infty.
\end{aligned}$$

Thus, $\mathbb{C}\mathrm{ov}(Y_{j+kd}, Y_{i+kd}) = \sigma_{\varepsilon,ij} / (1 - \alpha_i \alpha_j)$.

Let s be a multiple of d and let $s/d = m \in \mathbb{N}$ (the case $s/d = -m$, $m \in \mathbb{N}$, is similar).

Then $t - s = j + (k - m)d$ and

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(Y_t, Y_{t-s}) &= \mathbb{C}\mathrm{ov}(Y_{j+kd},\ Y_{j+(k-m)d}) \\
&= \mathbb{C}\mathrm{ov}\big(\alpha_j \circ Y_{j+(k-1)d} + \varepsilon_{j+kd},\ Y_{j+(k-m)d}\big) \\
&= \alpha_j \mathbb{C}\mathrm{ov}\left(\alpha_j \circ Y_{j+(k-2)d} + \varepsilon_{j+(k-1)d},\ Y_{j+(k-m)d}\right) \\
&= \ldots = \alpha_j^m \mathbb{C}\mathrm{ov}\left(Y_{j+(k-m)d}, Y_{j+(k-m)d}\right) \\
&= \alpha_j^m \frac{\sigma_{\varepsilon,jj} + \alpha_j \mu_{\varepsilon,j}}{1 - \alpha_j^2}.
\end{aligned}$$

If $s = 0$, then we have the variance given in 5.1.1 (c).

Let $t - s = i + md$, $m \in \mathbb{Z}$ and $i \in \{1, \ldots, d\}$ such that $i \neq j$ and $m \neq k$. Then, since $\mathbb{C}\mathrm{ov}(\varepsilon_{j+kd}, \varepsilon_{i+md}) = 0$ for all $m \neq k$, it can be shown that $\mathbb{C}\mathrm{ov}(Y_{j+kd}, Y_{i+md}) = 0$.

5.1.1 (e) The equality holds by eq. (5.1.1) and the properties of the binomial thinning operator.

D.2 Proof of Proposition 5.2.1

We begin by proving the stationarity of $\{\mathbf{Y}_k\}$. From the definition of a periodically correlated process (see Hurd and Miamee (2007, Def. 1.4)), we have that $\{Y_t\}$ in Proposition 5.1.1 (a) is periodically correlated with period d. Then, from Gladyshev (1961) it follows that process $\{\mathbf{Y}_k\}$, defined in eq. (5.2.1) is stationary (see also Hurd and Miamee (2007)). Furthermore, eqs. (5.2.1) have a unique stationary solution which is causal, when written as a VAR(1) process (see eq. (5.3.1)), since the roots of $\det(\mathbf{I}-\mathbf{A}z)=0$ are all outside the unit circle.

Next, we present proofs of the properties in Proposition 5.2.1.

5.2.1 (a) Using property 5.1.1 (b) and the fact that each row of (5.2.1) corresponds to an INAR process $Y_{j+kd}=\alpha_j \circ Y_{j+(k-1)d}+\varepsilon_{j+kd}$ verifies the first equality. Because the process mean vector is constant, taking the expectations of both sides of (5.2.1) we get

$$\boldsymbol{\mu}_{\mathbf{Y}} = (\mathbf{I}-\mathbf{A})^{-1}\boldsymbol{\mu}_{\varepsilon},$$

where $\boldsymbol{\mu}_{\varepsilon}=\mathbb{E}(\mathbf{Z}_k)$. Since $\mathbf{I}-\mathbf{A}$ is diagonal, calculating its inverse confirms the second equality.

5.2.1 (b) The proof is evident by applying the properties of the binomial thinning operator.

5.2.1 (c) The proof is obvious.

5.2.1 (d) The elements of $\boldsymbol{\Gamma}(0)$ can be calculated from Proposition 5.1.1 (c) and 5.1.1 (d):

$$\gamma_{ij}(0)=\mathbb{C}\mathrm{ov}(Y_{i+kd},Y_{j+kd})=\frac{\sigma_{\varepsilon,ij}+\alpha_i\mu_{\varepsilon,i}\mathbb{1}_{\{i=j\}}}{1-\alpha_i\alpha_j}.$$

Furthermore, applying the binomial thinning operator properties and the fact that ε_t are intra-seasonally dependent, $\boldsymbol{\Gamma}(h)$, $h>0$, can be expressed as

$$\mathbb{C}\mathrm{ov}(\mathbf{A}\circ\mathbf{Y}_{k-1}+\mathbf{Z}_k,\mathbf{Y}_{k-h})=\mathbf{A}\mathbb{C}\mathrm{ov}(\mathbf{Y}_{k-1},\mathbf{Y}_{k-h})=\cdots=\mathbf{A}^h\boldsymbol{\Gamma}(0),$$

which yields property 5.2.1 (d).

D.3 Proof of equation (5.3.2)

To express the multivariate INAR(1) process in eq. (5.2.1) as a VAR(1) process in eq. (5.3.1), we simply add and subtract the conditional mean, defined in Proposition 5.2.1(b), to get

$$\mathbf{Y}_k = \boldsymbol{\mu}_\varepsilon + \mathbf{A}\mathbf{Y}_{k-1} + \boldsymbol{e}_k,$$

where $\boldsymbol{e}_k = \mathbf{A} \circ \mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1} + \mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon$. Then, the expected value of $\boldsymbol{e}_k$ is

$$\mathbb{E}(\boldsymbol{e}_k) = \mathbf{A}\mathbb{E}(\mathbf{Y}_{k-1}) - \mathbf{A}\mathbb{E}(\mathbf{Y}_{k-1}) + \boldsymbol{\mu}_\varepsilon - \boldsymbol{\mu}_\varepsilon = 0$$

and the variance-covariance matrix of $\boldsymbol{e}_k$ is

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{e}} = \mathbb{E}\left(\boldsymbol{e}_k\boldsymbol{e}_k^\top\right) &= \mathbb{E}(\mathbf{A} \circ \mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1})(\mathbf{A} \circ \mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1})^\top \\ &+ \mathbb{E}(\mathbf{A} \circ \mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1})(\mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon)^\top \\ &+ \mathbb{E}(\mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon)(\mathbf{A} \circ \mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1})^\top + \mathbb{E}(\mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon)(\mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon)^\top .\end{aligned}$$

By applying the binomial thinning operator properties, the first term can be expressed as

$$\mathbb{E}(\mathbf{A} \circ \mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1})(\mathbf{A} \circ \mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1})^\top = \operatorname{diag}(\mathbf{B}\mathbb{E}(\mathbf{Y}_{k-1}))$$

and the remaining terms as

$$\mathbb{E}(\mathbf{A} \circ \mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1})(\mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon)^\top = \mathbf{0}$$

and

$$\mathbb{E}(\mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon)(\mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon)^\top = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}},$$

which reduces the covariance matrix expression to

$$\boldsymbol{\Sigma}_{\boldsymbol{e}} = \operatorname{diag}(\mathbf{B}\mathbb{E}(\mathbf{Y}_{k-1})) + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}.$$

It is then straightforward to verify that the final equality in eq. (5.3.2) holds true.

D.4 Proof of Proposition 5.3.1

It is straightforward to verify that equality $\boldsymbol{\beta} = \mathbf{R}\boldsymbol{\gamma}$ is satisfied provided restriction matrix $\mathbf{R}^\top = [\mathbf{R}_0^\top, \ldots, \mathbf{R}_{d-1}^\top]$ with $\mathbf{R}_k = (r_{i,j}^{(k)})$ given in (5.3.5), when the off-diagonal entries of $\mathbf{A}$ in eq. (5.2.1) are zeros.

Vectorizing eq. (5.3.3), we get

$$\operatorname{vec}(\mathbf{Y}) \;=\; \operatorname{vec}(\mathbf{X}\mathscr{B}\mathbf{I}_d) + \operatorname{vec}(\mathbf{E}) = (\mathbf{I}_d \otimes \mathbf{X})\,\mathbf{R}\boldsymbol{\gamma} + \operatorname{vec}(\mathbf{E}).$$

Noting that the covariance matrix of $\operatorname{vec}(\mathbf{E})$ is $\boldsymbol{\Sigma}_{\boldsymbol{e}} \otimes \mathbf{I}_n$, we want to choose an estimator of $\boldsymbol{\gamma}$, which minimizes

$$\begin{aligned} S(\boldsymbol{\gamma}) &= \operatorname{vec}(\mathbf{E})^\top (\boldsymbol{\Sigma}_{\boldsymbol{e}} \otimes \mathbf{I}_n)^{-1} \operatorname{vec}(\mathbf{E}) \qquad\qquad (\text{D.4.1}) \\ &= [\operatorname{vec}(\mathbf{Y}) - (\mathbf{I}_d \otimes \mathbf{X})\,\mathbf{R}\boldsymbol{\gamma}]^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{I}_n\right) [\operatorname{vec}(\mathbf{Y}) - (\mathbf{I}_d \otimes \mathbf{X})\,\mathbf{R}\boldsymbol{\gamma}]. \end{aligned}$$

Let $\mathscr{A} := (\mathbf{I}_d \otimes \mathbf{X})\,\mathbf{R}$ and $\boldsymbol{y} = \operatorname{vec}(\mathbf{Y})$. Then eq. (D.4.1) can be written as:

$$\begin{aligned} S(\boldsymbol{\gamma}) &= \boldsymbol{y}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{I}_n\right) \boldsymbol{y} - \boldsymbol{y}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{I}_n\right) \mathscr{A}\boldsymbol{\gamma} - \boldsymbol{\gamma}^\top \mathscr{A}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{I}_n\right) \boldsymbol{y} \\ &+ \boldsymbol{\gamma}^\top \mathscr{A}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{I}_n\right) \mathscr{A}\boldsymbol{\gamma} \\ &= \boldsymbol{y}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{I}_n\right) \boldsymbol{y} - \boldsymbol{y}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{X}\right) \mathbf{R}\boldsymbol{\gamma} - \boldsymbol{\gamma}^\top \mathbf{R}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{X}^\top\right) \boldsymbol{y} \\ &+ \boldsymbol{\gamma}^\top \mathbf{R}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \left(\mathbf{X}^\top \mathbf{X}\right)\right) \mathbf{R}\boldsymbol{\gamma}, \end{aligned}$$

which yields the REG-CLS estimate

$$\widehat{\gamma} = \left[\mathbf{R}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \left(\mathbf{X}^\top \mathbf{X}\right)\right) \mathbf{R}\right]^{-1} \mathbf{R}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{X}^\top\right) \boldsymbol{y}.$$

Following Lütkepohl (2007, p. 197) we substitute $\boldsymbol{\Sigma}_{\boldsymbol{e}}$ with its estimate, calculated using the unrestricted CLS estimates from eq. (5.3.4). This completes the proof.

D.5 Proof of Proposition 5.3.2

The proof is analogous to Lütkepohl (2007, Prop. 5.3). Noting that

$$\begin{aligned} \widehat{\boldsymbol{\gamma}} &= \left[\mathbf{R}^\top (\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes (\mathbf{X}^\top \mathbf{X}))\mathbf{R}\right]^{-1} \mathbf{R}^\top (\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{X}^\top) \left[(\mathbf{I}_d \otimes \mathbf{X})\mathbf{R}\boldsymbol{\gamma} + \operatorname{vec}(\mathbf{E})\right] \\ &= \boldsymbol{\gamma} + \left[\mathbf{R}^\top (\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes (\mathbf{X}^\top \mathbf{X}))\mathbf{R}\right]^{-1} \mathbf{R}^\top (\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{X}^\top) \operatorname{vec}(\mathbf{E}) \\ &= \boldsymbol{\gamma} + \left[\mathbf{R}^\top (\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes (\mathbf{X}^\top \mathbf{X}))\mathbf{R}\right]^{-1} \mathbf{R}^\top (\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1} \otimes \mathbf{I}_{d+1}) \operatorname{vec}(\mathbf{X}^\top \mathbf{E}), \end{aligned}$$

we have

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) \quad = \quad \left[\mathbf{R}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}\otimes\left(\frac{\mathbf{X}^\top\mathbf{X}}{n}\right)\right)\mathbf{R}\right]^{-1}\mathbf{R}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}\otimes\mathbf{I}_{d+1}\right)\frac{1}{\sqrt{n}}\mathrm{vec}\left(\mathbf{X}^\top\mathbf{E}\right).$$

Furthermore, from the assumptions of Proposition 5.3.2 and because $\mathbf{Z}_k$ are independent vectors, the conditions in Lütkepohl (2007, Prop. 5.3) are satisfied and we have that

$$\frac{\mathbf{X}^\top\mathbf{X}}{n}\xrightarrow{\mathrm{P}}\boldsymbol{\Gamma}_{\boldsymbol{X}}(0) \quad := \quad \mathbb{E}\left(\mathbf{X}_k^\top\mathbf{X}_k\right) \text{ as } n\to\infty$$

and

$$\frac{1}{\sqrt{n}}\mathrm{vec}\left(\mathbf{X}^\top\mathbf{E}\right)\xrightarrow{\mathrm{d}}\mathscr{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_{\boldsymbol{e}}\otimes\boldsymbol{\Gamma}_{\boldsymbol{X}}(0)\right) \text{ as } n\to\infty,$$

which yields that $\sqrt{n}\left(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)$ is asymptotically normal with covariance matrix

$$\begin{aligned}\left[\mathbf{R}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}\otimes\boldsymbol{\Gamma}_{\boldsymbol{X}}(0)\right)\mathbf{R}\right]^{-1}\mathbf{R}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}\otimes\mathbf{I}_{d+1}\right)\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}\otimes\boldsymbol{\Gamma}_{\boldsymbol{X}}(0)\right)\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}\otimes\mathbf{I}_{d+1}\right)&\\ \times\mathbf{R}\left[\mathbf{R}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}\otimes\boldsymbol{\Gamma}_{\boldsymbol{X}}(0)\right)\mathbf{R}\right]^{-1} = \left[\mathbf{R}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}\otimes\boldsymbol{\Gamma}_{\boldsymbol{X}}(0)\right)\mathbf{R}\right]^{-1}&.\end{aligned}$$

Vilnius University Press
9 Saulėtekio Ave., Building III, LT-10222 Vilnius
Email: info@leidykla.vu.lt,
www.leidykla.vu.lt
Print run 20