Copula based BINAR models with applications

Andrius Buteikis[*1]

[1]*Faculty of Mathematics and Informatics, Vilnius University, Naugarduko st. 24, LT-03225, Vilnius, Lithuania*

In this paper we study the problem of modelling the integer-valued vector observations. We consider the BINAR(1) models defined via copula-joint innovations. We review different parameter estimation methods and analyse estimation methods of the copula dependence parameter. We also examine the case where seasonality is present in integer-valued data and suggest a method of deseasonalizing them. Finally, an empirical application is carried out.

Keywords: Count data, BINAR, Poisson, Negative binomial distribution, Copula.

1 Introduction

Different financial institutions that issue loans do so following company-specific (and/or country-defined) rules which act as a safeguard so that loans are not issued to people who are known to be insolvent. 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 (for some examples of copula applications see [2], [4], [5] and [6]). While these studies were carried out for continuous data, there is less developed literature on discrete models created with copulas: [7] discussed the differences and challenges of using copulas for discrete data compared to continuous data. By using bivariate integer-valued autoregressive models (BINAR) it is possible to account for both the discreteness and autocorrelation of the data. Furthermore, copulas can be used to model the dependence of innovations in the BINAR(1) models: [9] used the Frank copula and normal copula to model the dependence of the innovations of the BINAR(1) model.

*Corresponding author: andrius.buteikis@mif.vu.lt

In this short paper we analyse different BINAR(1) model with copula-joint innovations parameter estimation methods. We also discuss some issues concerning the seasonality in integer-valued data and suggest a method of de-seasonalizing them. Finally, in order to analyse the presence of autocorrelation and copula dependence in loan data, an empirical application is carried out on weekly loan data. Estimation method comparisons and additional numerical results can be found in [3].

The paper is organized as follows. Section 2 presents the BINAR(1) process, Section 3 presents the definition of copulas. Section 4 compares different estimation methods for the BINAR(1) model. Seasonal adjustment of integer-valued data is presented in Section 5. In Section 6 an empirical application is carried out using different combinations of copula functions and marginal distribution functions. Conclusions are presented in Section 7.

2 The bivariate INAR(1) process

The BINAR(1) process was introduced in [11]. In this section we will provide the definition of the BINAR(1) model.

Definition 1. Let $\mathbf{R}_t = [R_{1,t}, R_{2,t}]'$, $t \in \mathbb{Z}$ be a sequence of independent identically distributed (i.i.d.) non-negative integer-valued bivariate random variables. A bivariate integer-valued autoregressive process of order 1 (BINAR(1)), $\mathbf{X}_t = [X_{1,t}, X_{2,t}]'$, $t \in \mathbb{Z}$, is defined as:

$$\mathbf{X}_t = \mathbf{A} \circ \mathbf{X}_{t-1} + \mathbf{R}_t = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} R_{1,t} \\ R_{2,t} \end{bmatrix}, \quad t \in \mathbb{Z}, \tag{1}$$

where $\alpha_j \in [0, 1)$, $j = 1, 2$, and the symbol '$\circ$' is the thinning operator which also acts as the matrix multiplication. We have that $\alpha_j \circ X_{j,t-1} := \sum_{i=1}^{X_{j,t-1}} Y_{j,t,i}$ and $Y_{j,t,1}, Y_{j,t,2}, \dots$ is a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(Y_{j,t,i} = 1) = \alpha_j = 1 - \mathbb{P}(Y_{j,t,i} = 0)$, $\alpha_j \in [0, 1)$, such that these sequences are mutually independent and independent of the sequence $\mathbf{R}_t$, $t \in \mathbb{Z}$. For each t, $\mathbf{R}_t$ is independent of $\mathbf{X}_s$, $s < t$.

A number of thinning operator properties are provided in [12] and [13]. Properties of the BINAR(1) model can be easily derived and a number of these are provided in [12]. We will expand on the work by [9] and [11] by analysing additional copulas for the BINAR(1) model innovation distribution as well as estimation methods for the distribution parameters.

3 Copulas

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. More information about Copula theory, properties and applications can be found in [10] and [8].

Since innovations of a BINAR(1) model are non-negative integer-valued random variables, one needs to consider copulas linking discrete distributions. According to Sklar's theorem [14], if F_1 and F_2 are discrete marginals then a unique copula representation exists only for values in the range of $\mathrm{Ran}(F_1) \times \mathrm{Ran}(F_2)$. However, the lack of uniqueness does not pose a problem in empirical applications because it implies that there may exist more than one copula which describes the distribution of the empirical data. Bivariate copulas which will be used when constructing and evaluating the BINAR(1) model in this paper are:

- The Farlie-Gumbel-Morgenstern (FGM) copula with $\theta \in [-1, 1]$:

$$C(u_1, u_2; \theta) = u_1 u_2 (1 + \theta(1 - u_1)(1 - u_2)),$$

- The Frank copula with $\theta \in (-\infty, \infty) \setminus \{0\}$:

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

where $u_1 := F_1(x_1)$, $u_2 := F_2(x_2)$. Here θ is the dependence parameter and F_1, F_2 - marginal cdfs. See [10] for properties of these copulas.

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

In this section we examine different BINAR(1) model parameter estimation methods. Let $\mathbf{X}_t = (X_{1,t}, X_{2,t})'$ be a non-negative integer-valued time series given in Def. 1, where the joint distribution of $(R_{1,t}, R_{2,t})'$, with marginals F_1, F_2, is linked by a copula $C(\cdot, \cdot)$: $\mathbb{P}(R_{1,t} \leq x_1, R_{2,t} \leq x_2) = C(F_1(x_1), F_2(x_2))$ and let $C(F_1(x_1), F_2(x_2)) = C(F_1(x_1), F_2(x_2); \theta)$, where θ is a dependence parameter.

4.1 Conditional least squares (CLS) estimation

The Conditional Least Squares (CLS) estimator minimizes the squared distance between $\mathbf{X}_t$ and its conditional expectation. Similarly to the method in [13] for the INAR(1) model, we construct the CLS estimator in the case of the BINAR(1) model. The CLS estimators of $\alpha_j, \lambda_j,\ j = 1, 2$ are found by minimizing the sum

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

The asymptotic properties of the CLS estimators for the INAR(1) model case are provided in [13]. Assume now that the Poisson innovations $R_{1,t}$ and $R_{2,t}$ with parameters λ_1 and λ_2, respectively, are joint by a copula with dependence parameter θ. In order to estimate θ, [3] minimized the sum of squared differences

$$S^{(M_1,M_2)} \quad = \quad \sum_{t=2}^{N} \left(\widetilde{X}_{1,t}^{\mathrm{CLS}} \widetilde{X}_{2,t}^{\mathrm{CLS}} - \gamma^{(M_1,M_2)}(\widehat{\lambda}_1^{\mathrm{CLS}}, \widehat{\lambda}_2^{\mathrm{CLS}}; \theta) \right)^2, \quad (3)$$

where

$$\widetilde{X}_{j,t}^{\mathrm{CLS}} \quad := \quad X_{j,t} - \widehat{\alpha}_j^{\mathrm{CLS}} X_{j,t-1} - \widehat{\lambda}_j^{\mathrm{CLS}}, \quad j = 1, 2,$$

$$\gamma^{(M_1,M_2)}(\lambda_1, \lambda_2; \theta) \quad := \quad \sum_{k=1}^{M_1} \sum_{l=1}^{M_2} kl\, c(F_1(k; \lambda_1), F_2(l; \lambda_2); \theta) - \lambda_1 \lambda_2,$$

where $c(F_1(k; \lambda_1), F_2(s; \lambda_2); \theta)$ is the joint probability mass function and M_1 and M_2 are used to approximate the covariance $\gamma(\lambda_1, \lambda_2; \theta)$ as described in [3].

4.2 Conditional maximum likelihood (CML) estimation

BINAR(1) models can also be estimated via conditional maximum likelihood (CML) (see [11] and [9]). The log conditional likelihood function is:

$$\ell \quad = \quad \sum_{t=2}^{N} \log \mathbb{P}(X_{1,t} = x_{1,t}, X_{2,t} = x_{2,t} | X_{1,t-1} = x_{1,t-1}, X_{2,t-1} = x_{2,t-1})$$

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

$$\ell(\alpha_1, \alpha_2, \lambda_1, \lambda_2, \theta) \longrightarrow \max_{\alpha_1, \alpha_2, \lambda_1, \lambda_2, \theta}. \quad (4)$$

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

For other marginal distribution cases where the marginal distribution has parameters other than λ_j, equation (4) would need to be minimized by those additional parameters.

4.3 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 λ_j are easily derived, [3] proposed to substitute the parameters of the marginal distributions in eq. (4) with CLS estimates from eq. (2). Then we would only need to maximize ℓ with respect to a single dependence parameter θ.

4.4 Estimation method comparison via Monte Carlo simulation

A Monte Carlo simulation was carried out in [3] in order to compare the estimation methods. The estimates of the dependence parameter were similar in terms of MSE and bias for both CML and Two-step estimation method.

5 Seasonality

Assume now that the nonnegative integer-valued time series can be written in the following form $\mathbf{Z}_t = \mathbf{S}_t + \mathbf{X}_t$, where $\mathbf{X}_t$ is defined by equation (1) and $\mathbf{S}_t = (S_{1,t}, S_{2,t})'$ is the (deterministic) integer-valued seasonal component with period d, where $S_{j,t} = S_{j,t+d}$, $\forall t$ and $j = 1, 2$ and $\sum_{k=0}^{d-1} S_{j,t+k} = 0$.

In order to remove the seasonal effect but keep the nonnegative, integer-valued properties of the data, we defined the operator $s(L) = 1 + L + ... + L^{d-1}$, where $L^k Z_t = Z_{t-k}$, $k \geq 0$. By applying this operator, the seasonal component is removed and the sample size decreases by $d-1$ observations. Alternatively, data can also be aggregated to a lower frequency (e.g. from daily to weekly data) in order to remove the seasonal effect at the cost of reducing the sample size d times. Finally, one can extend the seasonal INAR(1) model proposed in [1] to the BINAR(1) case.

Comparisons of these different seasonal adjustment methods is left for future research.

6 Application on default loan data

In this section we estimate a BINAR(1) model with the joint innovation distribution modelled by a copula cdf for empirical data. The dataset consists of weekly data on loans issued in Spain from October 21st, 2013, to January 1st, 2016 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 by considering a BINAR(1) model with different copulas for the innovations. For the marginal distributions of the innovations we considered Poisson as well as negative binomial distributions. We used the Two-step estimation method to estimate parameters. The dependence and variance parameter estimates when both marginals are negative binomial are provided in Table 1. Additional modelling results are provided in [3].

Table 1: Dependence and variance parameter estimates for BINAR(1) model via Two-step estimation method

Copula	$\hat{\theta}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	AIC
FGM	0.8927	6.5581	45.3683	1466.1542
	(0.1867)	(1.2402)	(7.5522)	
Frank	2.3848	6.5875	45.426	1466.9795
	(0.5337)	(1.2613)	(7.5774)	

Overall, both Frank and FGM copulas provide similar fit in terms of AIC, regardless of the selected marginal distributions. 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. Furthermore, the estimated dependence parameter is positive for the Frank and FGM copula cases, which indicates that there is a positive dependence between defaulted and non-defaulted loans.

7 Conclusions

In this short paper we have analysed different estimation methods for estimating parameters of a BINAR(1) model, including the dependence parameter of its innovations, which are linked via a copula. According to Monte Carlo simulations carried out in [3], BINAR(1) parameter estimates via CML had the smallest MSE and bias, however, estimates of the dependence parameter via CML and Two-step methods were similar. We also suggested a method to seasonally adjust the integer-valued data which exhibits a seasonal variation.

An empirical application on loan data was carried out and BINAR(1) models were estimated using different combinations of copula functions and marginal distribution functions. Additional estimation results are provided in [3]. The FGM copula provided the best model fit with Frank copula being very close in terms of AIC values. 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. Furthermore, the estimated copula dependence parameter indicates that the dependence between defaulted and non defaulted loans is positive.

Finally, one can apply different copula functions in order to analyse whether the loan data exhibits different forms of dependence. Lastly, the model can be extended by analysing the presence of structural changes within the data as well as extending the BINAR(1) model with copula joint innovations to account for the past values of other time series rather than only itself.

References

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

[2] D. Brigo and A. Pallavicini and R. Torresetti. *Credit Models and the Crisis: A Journey into CDOs, Copulas, Correlations and Dynamic Models.* Wiley, United Kingdom, 2011.

[3] A. Buteikis and R. Leipus. *Application of copula-based BINAR models in loan modelling.* Preprint.

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

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

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

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

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

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

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

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

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

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

[14] M. Sklar. *Fonctions de Repartition a n Dimensions et Leurs Marges.* Publications de l'Institut de Statistique de L'Universite de Paris, 8, 229–231, 1959.