CLADAG 2021

BOOK OF ABSTRACTS AND SHORT PAPERS

13th Scientific Meeting of the Classification and Data Analysis Group
Firenze, September 9-11, 2021

edited by
Giovanni C. Porzio
Carla Rampichini
Chiara Bocci

FIRENZE UNIVERSITY PRESS

PROCEEDINGS E REPORT
ISSN 2704-601X (PRINT) - ISSN 2704-5846 (ONLINE)

SCIENTIFIC PROGRAM COMMITTEE

Giovanni C. Porzio (chair) (University of Cassino and Southern Lazio - Italy)

Silvia Bianconcini (University of Bologna - Italy)
Christophe Biernacki (University of Lille - France)
Paula Brito (University of Porto - Portugal)
Francesca Marta Lilja Di Lascio (Free University of Bozen-Bolzano - Italy)
Marco Di Marzio ("Gabriele d'Annunzio" University of Chieti-Pescara - Italy)
Alessio Farcomeni ("Tor Vergata" University of Rome - Italy)
Luca Frigau (University of Cagliari - Italy)
Luis Ángel García Escudero (University of Valladolid - Spain)
Bettina Grün (Vienna University of Economics and Business - Austria)
Salvatore Ingrassia (University of Catania - Italy)
Volodymyr Melnykov (University of Alabama - USA)
Brendan Murphy (University College Dublin -Ireland)
Maria Lucia Parrella (University of Salerno - Italy)
Carla Rampichini (University of Florence - Italy)
Monia Ranalli (Sapienza University of Rome - Italy)
J. Sunil Rao (University of Miami - USA)
Marco Riani (University of di Parma - Italy)
Nicola Salvati (University of Pisa - Italy)
Laura Maria Sangalli (Polytechnic University of Milan - Italy)
Bruno Scarpa (University of Padua - Italy)
Mariangela Sciandra (University of Palermo - Italy)
Luca Scrucca (University of Perugia - Italy)
Domenico Vistocco (Federico II University of Naples - Italy)
Mariangela Zenga (University of Milan-Bicocca - Italy)

LOCAL PROGRAM COMMITTEE

Carla Rampichini (chair) (University of Florence - Italy)

Chiara Bocci (University of Florence - Italy)
Anna Gottard (University of Florence - Italy)
Leonardo Grilli (University of Florence - Italy)
Monia Lupparelli (University of Florence - Italy)
Maria Francesca Marino (University of Florence - Italy)
Agnese Panzera (University of Florence - Italy)
Emilia Rocco (University of Florence - Italy)
Domenico Vistocco (Federico II University of Naples - Italy)

CLADAG 2021 BOOK OF ABSTRACTS AND SHORT PAPERS

13th Scientific Meeting of the Classification and Data Analysis Group
Firenze, September 9-11, 2021

edited by
Giovanni C. Porzio
Carla Rampichini
Chiara Bocci

FIRENZE UNIVERSITY PRESS
2021

CLADAG 2021 BOOK OF ABSTRACTS AND SHORT PAPERS : 13th Scientific Meeting of the Classification and Data Analysis Group Firenze, September 9-11, 2021/ edited by Giovanni C. Porzio, Carla Rampichini, Chiara Bocci. — Firenze : Firenze University Press, 2021.
(Proceedings e report ; 128)

https://www.fupress.com/isbn/9788855183406

ISSN 2704-601X (print)
ISSN 2704-5846 (online)
ISBN 978-88-5518-340-6 (PDF)
ISBN 978-88-5518-341-3 (XML)
DOI 10.36253/978-88-5518-340-6

Graphic design: Alberto Pizarro Fernández, Lettera Meccanica SRLs
Front cover: Illustration of the statue by Giambologna, *Appennino* (1579-1580) by Anna Gottard

CLAssification and Data
Analysis Group (CLADAG)
of the Italian Statistical
Society (SIS)

FUP Best Practice in Scholarly Publishing (DOI https://doi.org/10.36253/fup_best_practice)
All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Boards of the series. The works published are evaluated and approved by the Editorial Board of the publishing house, and must be compliant with the Peer review policy, the Open Access, Copyright and Licensing policy and the Publication Ethics and Complaint policy.

Published by Firenze University Press
Firenze University Press
Università degli Studi di Firenze
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

This book is printed on acid-free paper
Printed in Italy

INDEX

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Giovanni C. Porzio, Carla Rampichini, Chiara Bocci (edited by), *CLADAG 2021 Book of abstracts and short papers. 13th Scientific Meeting of the Classification and Data Analysis Group Firenze, September 9-11, 2021*, © 2021 Author(s), content CC BY 4.0 International, metadata CC0 1.0 Universal, published by Firenze University Press (www.fupress.com), ISSN 2704-5846 (online), ISBN 978-88-5518-340-6 (PDF), DOI 10.36253/978-88-5518-340-6

Hidden Markov and regime switching copula models for state allocation in multiple time-series

Francesco Bartolucci[1], Fulvia Pennoni[2], and Federico P. Cortese[3]

[1] Department of Economics, University of Perugia
(e-mail: francesco.bartolucci@unipg.it)

[2] Department of Statistics and Quantitative Methods, University of Milano-Bicocca
(e-mail: fulvia.pennoni@unimib.it)

[3] Department of Economics, Management and Statistics, University of Milano-Bicocca
(e-mail: f.cortese5@campus.unimib.it)

ABSTRACT: We consider hidden Markov and regime-switching copula models as approaches for state allocation in multiple time-series, where state allocation means prediction of the latent state characterizing each time occasion based on the observed data. This dynamic clustering, performed under the two model specifications, takes the correlation structure of the time-series into account. Maximum likelihood estimation of the model parameters is carried out by the expectation-maximization algorithm. For illustration we use data on the market of cryptocurrencies characterized by periods of high turbulence in which interdependence among assets is marked.

KEYWORDS: daily log-returns, expectation-maximization algorithm, forecast, latent variables, model-based clustering

1 Introduction

In the analysis of multiple time-series, state allocation, namely prediction of the state or regime underlying the observed data at a certain time occasion, is an important task, especially in finance and related fields. This type of clustering is dynamic because a different state may be predicted at every time occasion and may be based on models representing each time-specific state by a discrete latent variable assuming, typically, a few possible values. In this contribution, we compare two different model specifications of this type: multivariate hidden Markov (HM) models (Zucchini *et al.*, 2017) and regime-switching (RS) copulas (Rodriguez, 2007).

Among HM models we consider, in particular, those based on the assumption that the time-specific vector of observable variables follows a conditional Gaussian distribution with parameters depending on the latent state.

RS copulas are instead based on a copula function, which may be chosen among the Clayton, the Gumbel, the Gaussian, or the Student-t, with parameters governed by a hidden Markov process of first-order so as to flexibly account for the correlation patterns between each pair of series.

The expectation-maximization (EM) algorithm (Dempster *et al.*, 1977) is used for maximum likelihood estimation of the parameters of both models. Model selection is performed to choose the most appropriate number of hidden states and evaluate the level of chain homogeneity over time (Bartolucci *et al.*, 2013). For the HM model, this selection is based on the Bayesian Information Criterion (BIC), and for RS copulas, it is also based on a goodness-of-fit procedure relying on the Cramér-von Mises statistic.

As an illustration we consider the problem of state allocation in analyzing time-series of the main cryptocurrencies daily log-returns over a three-year period.

2 Hidden Markov and Regime-Switching Copula Models

Let $\boldsymbol{y}_t$, $t = 1,2,\ldots$, be the vector where each element y_{tj}, $j = 1,\ldots,r$, corresponds to the value of time-series j at time occasion t, with r denoting the number of time-series under consideration. The main assumption of the multivariate HM model is that the random vectors $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ are conditionally independent given a hidden process $u_1, u_2, \ldots$ that follows a first-order Markov chain with k states, labeled from 1 to k. This process is governed by the initial probabilities $\pi_u = p(u_1 = u)$, $u = 1,\ldots,k$, and the transition probabilities $\pi_{u|\bar{u}} = p(u_t = u|u_{t-1} = \bar{u})$, $t = 2,\ldots,$ $\bar{u}, u = 1,\ldots,k$. We assume a Gaussian distribution for the observations at every time occasion, that is, $\boldsymbol{y}_t \mid u_t = u \sim N_r(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)$, where $\boldsymbol{\mu}_u$ and $\boldsymbol{\Sigma}_u$ are the mean vector and variance-covariance matrix for latent state u. The above assumptions imply that the conditional distribution of the time-series $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots,$ given the sequence of hidden states, may be expressed as $f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \mid u_1, u_2, \ldots) = \prod_t \phi(\boldsymbol{y}_t; \boldsymbol{\mu}_{u_t}, \boldsymbol{\Sigma}_{u_t})$, where $\phi(\cdot;\cdot)$ denotes the density of the multivariate Gaussian distribution. The manifest distribution of the multiple time-series has the following density function:

$$f(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots) = \sum_{u_1} \pi_{u_1} \phi(\boldsymbol{y}_1; \boldsymbol{\mu}_{u_1}, \boldsymbol{\Sigma}_{u_1}) \sum_{u_2} \pi_{u_2|u_1} \phi(\boldsymbol{y}_2; \boldsymbol{\mu}_{u_2}, \boldsymbol{\Sigma}_{u_2}) \cdots .$$

Concerning the copula model, we first consider only the bivariate case, so we define $\boldsymbol{y}_t = (y_{t1}, y_{t2})$ as a vector with elements y_{tj}, $j = 1,2$, corresponding to the observation for time-series j at time $t = 1,2,\ldots$ and F_1 and F_2 as the marginal cdfs of each time-series. Sklar's theorem (Sklar, 1959) allows us to separate the fitting of the marginal cdfs from the fitting of the joint distribution, represented by a copula function. This approach consists in estimating the two marginal distributions, obtaining $\hat{F}_1$ and $\hat{F}_2$, and then computing the normalized ranks of the pseudo-observations $\tilde{\boldsymbol{e}}_t = (\tilde{e}_{t1}, \tilde{e}_{t2})$ as $\tilde{e}_{tj} = \text{rank}(\hat{z}_{tj})/(T+1)$, with $\hat{z}_{tj} = \hat{F}_j(y_{tj})$, and T being the number of observed time occasions. Finally, for the pseudo-observations $\tilde{\boldsymbol{e}}_t$, an RS copula model is assumed based on a hidden homogeneous Markov process denoted as $v_1, v_2, \ldots$, with k states. The copula density indicated with $c(\cdot;\cdot)$ may be chosen among the Clayton, the Gumbel, the Gaussian, or the Student-t copulas, with state-specific parameter β_v. The density of the pseudo-observations is given by

$$f(\tilde{\boldsymbol{e}}_1, \tilde{\boldsymbol{e}}_2, \ldots) = \sum_{v_1} \pi_{v_1} c(\tilde{\boldsymbol{e}}_1; \beta_{v_1}) \sum_{v_2} \pi_{v_2|v_1} c(\tilde{\boldsymbol{e}}_2; \beta_{v_2}) \cdots,$$

and it is based on the initial and transition probabilities defined as above.

Given that the state sequence is not observable, a full maximum likelihood approach for estimating the parameters of both models is carried out through the EM algorithm. Following the current literature, model selection for the HM model is based on the BIC, and for the RS copula it is also performed through a goodness-of-fit procedure consisting in calculating a p-value referred to the Cramér-von Mises statistic for the hypothesis of correct model specification.

We compare the performance of HM models and RS copulas focusing on the crucial aspect of state allocation. The optimal state allocation is performed by finding the optimal joint sequence $\tilde{u}_1, \tilde{u}_2, \ldots$ (or $\tilde{v}_1, \tilde{v}_2, \ldots$) of unknown states given the corresponding observations. This clustering procedure, also known as global decoding, is achieved through the Viterbi algorithm (Viterbi, 1967), which is a dynamic programming algorithm.

We also aim at extending the RS copula approach to an arbitrary number of time-series r rather than to only 2. In this regard, we propose the composite likelihood approach (Varin *et al.*, 2011) for estimation, which is based on considering all possible ordered pairs of time-series among the available ones.

3 Application

As an illustration, for the HM model we consider the joint daily log-returns* of the five cryptocurrencies Bitcoin, Ethereum, Ripple, Litecoin, and Bitcoin

*provided by the Crypto Asset Lab: `https://cryptoassetlab.diseade.unimib.it/`.

Cash, for the period 2017-2020. For the RS copulas, allowing only for bivariate associations, we define four copulas where the bivariate vector of observations consists of the Bitcoin and each of the other four cryptocurrencies. Results for the HM model show that the minimum value of the BIC is reached considering a five-state heteroschedastic structure. According to these estimates, there are three negative regimes (in terms of estimated expected log-returns), with relatively high and positive correlations of Bitcoin with all the other cryptocurrencies, and two states with positive returns and lower correlations. Regarding the global decoding, these two states are the most likely in the first year of observation, and the other three states characterize the last two years.

Concerning the RS copulas, and considering as an example the couple of cryptocurrencies Bitcoin-Ethereum, we observe that a three-regime Clayton copula provides the best fit. Given that the Clayton copula allows for explicit computation of the lower tail correlation index, we estimate that two regimes provide zero or low values for the lower tail index, and the third regime provides high values for it. According to the optimal state sequence, we estimate that there is substantial interchangeability between the first two states in the whole period, whereas the third state is the most likely for the last year of observation.

References

BARTOLUCCI, F., FARCOMENI, A., & PENNONI, F. 2013. *Latent Markov Models for Longitudinal Data.* Boca Raton, FL: Chapman & Hall/CRC.

DEMPSTER, A. P., LAIRD, N. M., & RUBIN, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

RODRIGUEZ, J. C. 2007. Measuring financial contagion: A copula approach. *Journal of Empirical Finance*, **14**, 401–423.

SKLAR, M. 1959. Fonctions de repartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, **8**, 229–231.

VARIN, C., REID, N., & FIRTH, D. 2011. An overview of composite likelihood methods. *Statistica Sinica*, **21**, 5–42.

VITERBI, A. 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, **13**, 260–269.

ZUCCHINI, W., MACDONALD, I. L., & LANGROCK, R. 2017. *Hidden Markov Models for Time Series: An Introduction Using R.* Boca Raton, FL: CRC.