



HAL
open science

Three-stage estimation method for non-linear multiple time-series

Dominique Guegan, Giovanni de Luca, Giorgia Riviuccio

► **To cite this version:**

Dominique Guegan, Giovanni de Luca, Giorgia Riviuccio. Three-stage estimation method for non-linear multiple time-series. 2017. halshs-01439860

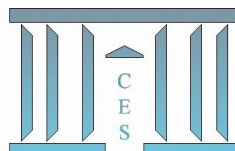
HAL Id: halshs-01439860

<https://shs.hal.science/halshs-01439860>

Submitted on 18 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Three-stage estimation method for non-linear
multiple time-series**

Dominique GUEGAN, Giovanni DE LUCA, Giorgia RIVIECCIO

2017.01



Three-stage estimation method for non-linear multiple time-series

Dominique Guegan* Giovanni De Luca[†] Giorgia Riviuccio[‡]

Abstract

We present the three-stage pseudo maximum likelihood estimation in order to reduce the computational burdens when a copula-based model is applied to multiple time-series in high dimensions. The method is applied to general stationary Markov time series, under some assumptions which include a time-invariant copula as well as marginal distributions, extending the results of Yi and Liao [2010]. We explore, via simulated and real data, the performance of the model compared to the classical vectorial autoregressive model, giving the implications

*Université Paris 1 Panthéon-Sorbonne, CES UMR 8174, 106 boulevard de l'Hopital 75647, Paris Cedex 13, France, phone: +33144078298, e-mail: dguegan@univ-paris1.fr, Labex Refi: this work has been achieved through the Laboratory of Excellence on Financial Regulation supported by PRES HeSam under the reference ANR-10-LABEX-0095

[†]Parthenope University of Naples, via Generale Parisi 38, 80122 Naples, Italy, phone: +39 0815474516, e-mail: giovanni.deluca@uniparthenope.it

[‡]Parthenope University of Naples, via Generale Parisi 38, 80122 Naples, Italy, phone: +39 0815474988, e-mail: giorgia.riviuccio@uniparthenope.it

of misspecified assumptions for margins and/or joint distribution and providing tail dependence measures of economic variables involved in the analysis.

Keywords

Copula function, Three stage estimator, Multiple time series.

1 Introduction

The 2008 financial crisis, concretized in a financial and economic default contagion, showed the inadequacy of linear co-movement measures, focusing the attention of researchers on nonlinear and asymmetric dependence structures, which take into account extreme value dependence (see, Longin and Solnik [2001], Pickands [1981], Dobrić and Schmid [2005], Lo and MacKinlay [1990]). Besides, time dependence analysis cannot be disregarded, especially given two pillars of dependence relationships, serial dependence and contemporaneous dependence, defined, respectively, as temporal auto-dependence of univariate time series and cross-dependence among contemporary variables possibly conditioned to a common set of their lagged values (see Joe [1997], Chen and Fan [2006], Abegaz and Naik-Nimbalkar [2007]).

Vector AutoRegressive (VAR) models (Hamilton [1994]) are able to capture the dynamic behavior of multiple time series, but their use is often limited in a linear and symmetric framework when a Gaussian distribution is adopted (see, e.g, DeMiguel et al. [2014]). In addition, a wrong specification of the joint distribution of the variables can induce a distorted evaluation of risk, especially due to

extreme events. In order to describe the serial and cross-sectional dependence among time series, copulas can be a practical method (Bianchi et al. [2010] and Brechmann and Czado [2015]). A copula function is a statistical tool that allows to consider all the information about the dependence structure between the components of a random vector. The major advantage of their adoption is the flexible specification of the marginal distributions, which do not necessarily belong to the same family, even if it pays for long and complicated parameter estimation procedures in multivariate context, when parametric approaches are followed, particularly in high dimensions. A common procedure among copula practitioners, when dealing with high-dimensional data analysis and maximum likelihood methods of parameter estimation, is to split the likelihood function in two parts, one for the margin parameters and the other for the dependence structure. This procedure is also known as Inference For Margins method (IFM) (Joe and Xu [1996]).

The aim of the paper is to provide an estimation procedure able to solve several problems in modelling multiple time series: (i) the separate estimate of two classes of dependence for multiple time series vectors; (ii) the description of a non linear serial and cross-sectional dependence structure for multivariate series; (iii) the measure of dependence among extreme values of the variables. In order to reach these objectives, we extend the results of Yi and Liao [2010] who have introduced a three stage estimator (3SPMLE), based on copula function, in order to evaluate separately these two classes of dependence relationships for time series vectors, serial and contemporaneous dependencies, in addition

to marginal parameters. While they have considered a non linear dependence structure limiting to the context of the analysis of the bivariate case of stationary Markov chain models, we estimate a non-linear VAR model based on copula functions in a multivariate context, which results more efficient of the classical VAR approach and allows to consider tail dependence estimates. In literature there exists an alternative, but competitive, approach defined COPAR model (Brechmann and Czado [2015]) which is based on vine copulas. In the paper we do not intend to compare our model with this approach because we want to give a different way to build non linear dependence structures for all copula practitioners, not only for vine ones. This comparison can be objective of future researches. To this end, we extend the three-stage pseudo maximum likelihood estimation of Yi and Liao [2010] in order to alleviate the computational burden when copula-based VAR model is applied in high dimensions. The method is applied to stationary Markov time series, under some assumptions which include a time-invariant copula as well as marginal distributions.

We explore, via simulated and real data, the performance of the model compared to the classical VAR, giving the implications of misspecified assumptions for margins and joint distribution and providing tail dependence measures of economic variables involved in the analysis. The paper is organized as follows. In Section 2 we introduce some theoretical backgrounds of copulas and VAR models, focusing then on the three stage estimation technique and giving a brief definition of tail dependence and its measure. In Section 3 we show the performance of the model on simulated data. Section 4 presents an empirical

analysis on MSCI world sector indices. Section 5 concludes.

2 Theoretical background: copula, VAR model and three stage estimation method

In this section, we give the main tools useful for measuring dependence: in the beginning we provide some theoretical background on copulas, then we discuss one of the classical models for multiple time serie, the Vectorial Autoregressive (VAR) model, in both a linear and a nonlinear framework, finally we show the three stage estimation approach, easily adaptable in the latter context.

Copula functions have been extensively used in the latest years. One of the point of strength of the copula function is the flexibility in describing multivariate data combined to a sufficiently ease of estimation. A p -dimensional copula function C is formally defined as a function of p variables u_1, \dots, u_p , each defined in $[0, 1]$ such that its range is the unit interval $[0, 1]$. Given a $p \times 1$ random vector Y with joint and marginal distribution functions, respectively, F and F_i ($i = 1, \dots, p$), according to the Sklar's theorem, the joint distribution function F can be expressed through a copula function,

$$F(y_1, \dots, y_p) = C(F_1(y_1), \dots, F_p(y_p))$$

where $F_i(y_i) = P(Y_i \leq y_i) = u_i$ is a vector of uniform variables in $[0, 1]^p$ (for $i = 1, \dots, p$). If F_i is continuous, then the copula C is unique; otherwise, C is uniquely determined on $[\text{rank}(F_1), \dots, \text{rank}(F_p)]$.

In a statistical perspective, the copula function is a function joining the

univariate distribution functions. As a result, the estimation of the joint distribution function can be obtained after specifying the univariate distributions and a function which determines the dependence structure.

If the joint distribution function is p -times differentiable, then

$$f(y_1, \dots, y_p; \boldsymbol{\theta}) = \prod_{i=1}^p f_i(y_i | \boldsymbol{\theta}_M) c(F_1(y_1), \dots, F_p(y_p); \boldsymbol{\theta}_C)$$

where $c(\cdot)$ is the copula density and $\boldsymbol{\theta} = [\boldsymbol{\theta}_M, \boldsymbol{\theta}_C]'$ is the parameter vector to estimate.

For this purpose, the log-likelihood at time t is

$$\ell_t(\boldsymbol{\theta}) = \sum_{i=1}^p \log f_i(y_i; \boldsymbol{\theta}_M) + \log c(F_1(y_1), \dots, F_p(y_p); \boldsymbol{\theta}_C) \quad (1)$$

The exact maximum likelihood method consists of maximizing the log-likelihood function

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^T \ell_t(\boldsymbol{\theta}).$$

However, the exact maximum likelihood method can be computationally very intensive, especially if p is large. The popular Inference for Margins (IFM) method reduces the computational burden assuming two steps, each associated to one of the two terms of the log-likelihood (1). In the first step, the p univariate distributions are estimated, such that the estimated distribution functions are inserted as arguments in the copula density. In the second step the copula density is maximized and the parameters describing the dependence structure are estimated.

Under regularity conditions (Cherubini et al. [2004]) the IFM estimator is consistent and asymptotically efficient and verifies the property of asymptoti-

cally normality

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}) \rightarrow (0, \mathcal{G}(\boldsymbol{\theta})^{-1})$$

where $\mathcal{G}(\boldsymbol{\theta})$ is the Godambe information matrix (see inter alia Nelsen [2006]).

The split of the maximization problem is certainly a winner strategy. It has been proposed in a context of p identically and independently distributed (iid) variables, whereas the objectives of the estimation problem are the parameters of the marginal distribution and the parameters of the dependence structure (auto-dependence or co-dependence), without allowing the coexistence of both of them. In order to take into account both types of dynamic dependence of multiple time series, auto-dependence and co-dependence, we have extended the estimator proposed by Yi and Liao [2010] considering more than two stationary first-order Markov chains processes and applying it to a general non-linear Vector Autoregressive (VAR) model which takes into account the serial and cross-dependence among the variables. A p -dimensional non-linear VAR(1) model which involves the function $g_i(\cdot)$ to build a non-linear temporal relationship among variables can be written as

$$\begin{aligned} y_{1,t} &= g_1(y_{1,t-1}, \dots, y_{p,t-1}) + \epsilon_{1t} \\ y_{2,t} &= g_2(y_{1,t-1}, \dots, y_{p,t-1}) + \epsilon_{2t} \\ \dots &= \dots \\ y_{p,t} &= g_p(y_{1,t-1}, \dots, y_{p,t-1}) + \epsilon_{pt} \end{aligned} \tag{2}$$

with $\boldsymbol{\epsilon}_t = [\epsilon_{1t}, \dots, \epsilon_{pt}]' \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$.

A special form of this general setup, the linear VAR (Hamilton [1994]), is

a very popular model, widely used in economics, but distinguished by two features: the linearity of the relationship among the p variables and the conditional Gaussian distribution for each variable.

Then, we assume that the conditional distribution will be described by a conditional copula function whose parameters will be estimated extending the three stage approach of Yi and Liao [2010].

We describe now, in details, the procedure for a bivariate nonlinear VAR(1) corresponding to $p = 2$ in (2).

Then, defined $z_{t-1} = [y_{1t-1}, y_{2t-1}]$, we get

$$F(y_{1t}|z_{t-1}, y_{2t}|z_{t-1}) = C(F(y_{1t}|z_{t-1}), F(y_{2t}|z_{t-1}))$$

where

$$F(y_{it}|z_{t-1}) = \frac{\partial^2 C(F(y_{it}), F(y_{1t-1}), F(y_{2t-1}))}{\partial F(y_{1t-1}) \partial F(y_{2t-1})} \quad \text{for } i = 1, 2$$

So

$$\begin{aligned} f(y_{1t}|z_{t-1}, y_{2t}|z_{t-1}) &= c(F(y_{1t}|z_{t-1}), F(y_{2t}|z_{t-1})) \cdot f(y_{1t}|z_{t-1}) \cdot f(y_{2t}|z_{t-1}) \\ &= c(F(y_{1t}|z_{t-1}), F(y_{2t}|z_{t-1})) \cdot \frac{f(y_{1t}, y_{1t-1}, y_{2t-1})}{f(y_{1t-1}, y_{2t-1})} \cdot \frac{f(y_{2t}, y_{1t-1}, y_{2t-1})}{f(y_{1t-1}, y_{2t-1})} \\ &= c(F(y_{1t}|z_{t-1}), F(y_{2t}|z_{t-1})) \cdot \frac{c(F(y_{1t}), F(y_{1t-1}), F(y_{2t-1}))}{c(F(y_{1t-1}), F(y_{2t-1}))} \\ &\quad \cdot \frac{c(F(y_{2t}), F(y_{1t-1}), F(y_{2t-1}))}{c(F(y_{1t-1}), F(y_{2t-1}))} \cdot f(y_{1t}) \cdot f(y_{2t}) \end{aligned}$$

It is now easy to obtain the likelihood as a product of all conditional densities, then the log-likelihood is composed of three parts which can be maximized separately in a hierarchical way:

$$\ell = \sum_{i=1}^2 \sum_{t=1}^T \ln f(y_{it})$$

$$\begin{aligned}
& + \sum_{i=1}^2 \sum_{t=2}^T \ln \frac{c(F(y_{it}), F(y_{1t-1}), F(y_{2t-1}))}{c(F(y_{1t-1}), F(y_{2t-1}))} \\
& + \sum_{t=2}^T \ln c(F(y_{1t}|z_{t-1}), F(y_{2t}|z_{t-1}))
\end{aligned}$$

The procedure starts with the estimation of the marginal distributions, which are then used to estimate the distribution function to be used as input in the second stage. Finally, in the third stage the conditional contemporaneous dependence is estimated. The copula functions used in the second and third stage do not need to be the same, so a high degree of flexibility is ensured.

It is straightforward to extend the formulation in the most general case of p variables and d lags. Defined $F(z_{t-k}) = [F(y_{1t-k}), \dots, F(y_{pt-k})](\forall k = 1, \dots, d)$, we have

$$\begin{aligned}
\ell & = \sum_{i=1}^p \sum_{t=1}^T \ln f(y_{it}) \\
& + \sum_{i=1}^p \sum_{t=d+1}^T \ln \frac{c(F(y_{it}), F(z_{t-1}), \dots, F(z_{t-d}))}{c(F(z_{t-1}), \dots, F(z_{t-d}))} \\
& + \sum_{t=d+1}^T \ln c(F(y_{1t}|z_{t-1}, \dots, z_{t-d}), \dots, F(y_{pt}|z_{t-1}, \dots, z_{t-d}))
\end{aligned}$$

The fitting of a copula function for the p -variate conditional distribution of $\mathbf{y}_t = [y_{1t}, \dots, y_{pt}]'$ can involve the presence of tail dependence between any bivariate subset of \mathbf{y}_t . Tail dependence is a useful copula-based measure which defines the relationship among extreme values. This measure of concordance is different for each family of copulas: it can be symmetric or asymmetric, or also absent (e.g. the Gaussian copula). The lower tail dependence coefficient, denoted as λ_L , is given by the probability of very low values of a variable, given

that very low values have occurred to the other variable. In terms of copula,

$$\lambda_L = \lim_{v \rightarrow 0^+} \frac{C(v, v)}{v}.$$

Analogously, the upper tail dependence coefficient is the probability of extremely high values of a variable, given that extremely high values have occurred to the other variable. In terms of copula,

$$\lambda_U = \lim_{v \rightarrow 1^-} \frac{1 - 2v + C(v, v)}{1 - v}.$$

The most important elliptical copulas, the Gaussian and Student's- t copula, present, respectively, null and symmetric tail dependence. This implies that the use of a Gaussian copula prevents from quantifying an association measure among extreme values, while the Student's- t copula allows such a relationship but with the constraint of admitting no difference between the two tails. In particular, for a bivariate Student's t copula with correlation ρ and degrees of freedom ν the tail dependence coefficients are coincident (Demarta and McNeil [2005]),

$$\lambda_L = \lambda_U = \lambda = 2t_{\nu+1} \left(-\sqrt{\nu+1} \sqrt{\frac{1-\rho}{1+\rho}} \right). \quad (3)$$

In general, a wrong specification of margins and/or joint distribution, which is quite common in Gaussian models, could result in an underestimation of risks related to extreme events. Multivariate lower and upper tail dependence coefficients are defined in De Luca and Riveccio [2012].

3 Simulation

In order to evaluate the impact of the use of the copula-based methodology comparing with a classical VAR modelling, we have carried out a simulation exercise. We have simulated $B = 1000$ samples of size $T = 1000$ assuming as data generating process (DGP) a VAR(1) model with 4 variables,

$$\begin{aligned}y_{1,t} &= 0.15y_{1,t-1} + 0.20y_{2,t-1} + 0.15y_{3,t-1} + 0.25y_{4,t-1} + \epsilon_{1t} \\y_{2,t} &= 0.13y_{1,t-1} + 0.20y_{2,t-1} + 0.15y_{3,t-1} + 0.25y_{4,t-1} + \epsilon_{2t} \\y_{3,t} &= 0.23y_{1,t-1} + 0.20y_{2,t-1} + 0.15y_{3,t-1} + 0.25y_{4,t-1} + \epsilon_{3t} \\y_{4,t} &= 0.33y_{1,t-1} + 0.20y_{2,t-1} + 0.15y_{3,t-1} + 0.25y_{4,t-1} + \epsilon_{4t}\end{aligned}$$

with the innovation vector $\epsilon_t = [\epsilon_{1t}\epsilon_{2t}\epsilon_{3t}\epsilon_{4t}]'$ distributed as Student's- t with degrees of freedom parameter $\nu = 5$ and the following variance-covariance matrix

$$\Sigma = \begin{bmatrix} 1.000 & 0.175 & 0.165 & 0.200 \\ 0.175 & 1.000 & 0.110 & 0.205 \\ 0.165 & 0.110 & 1.000 & 0.175 \\ 0.200 & 0.205 & 0.175 & 1.000 \end{bmatrix}$$

The assumption of quadrivariate Student's t implies that the univariate distribution of $y_{i,t}|\mathcal{I}_{t-1}$ ($i = 1, \dots, 4$) is also Student's t with the same degrees of freedom parameter ν (\mathcal{I}_{t-1} denotes the information at time $t - 1$). Then, we have estimated:

1. the Gaussian VAR(1) as benchmark model;
2. the copula-based VAR(1) model exploiting the three-stage approach.

We have then compared the results in terms of Akaike Information Criterion (AIC) and tail dependence coefficients. To implement the three stage approach, we adopt the following strategy

1. I stage: the marginal distributions are estimated as Student's t distributions;
2. II stage: the serial and cross-dependence are estimated using the Student's t copulas;
3. III stage: another Student's t copula is estimated for the conditional contemporaneous dependence structure.

In particular, the average values of the correlation coefficients of the third-stage Student's t copula are reported in $\bar{\hat{\Sigma}}$ (the histograms for each coefficient are reported in Figure 1),

$$\bar{\hat{\Sigma}} = \begin{bmatrix} 1.000 & 0.179 & 0.168 & 0.205 \\ 0.179 & 1.000 & 0.111 & 0.210 \\ 0.168 & 0.111 & 1.000 & 0.178 \\ 0.205 & 0.210 & 0.178 & 1.000 \end{bmatrix}$$

while the average value of the degrees of freedom parameter is 6.85 (see the histogram in Figure 2). The comparison of the copula approach with the Gaussian approach in terms of AIC has produced a strong result in favor of the copula summarized in Figure 3. The Akaike Information Criterion is almost smaller for the model estimated using the three-stage methodology, than is using copula functions. Moreover, the estimation of a Gaussian model implies no association

between extreme values, that is the nullity of the tail dependence coefficients. However, both the tail dependence coefficients for a bivariate Student's- t copula with correlation ρ and degrees of freedom ν are given in (3). So, according to the DGP, we estimate the tail dependence coefficients for the simulated data and the average tail dependence coefficients for the three-stage copula-based VAR which are positive even if slightly biased (the histograms for each tail dependence coefficient are reported in Figure 4). Table 1 reports the estimate results for both tail dependence types. This simple simulation exercise shows that an estimation approach divided into several stages allows to capture in a more efficient way the distributional features of a set of multivariate data. Differently, the use of a Gaussian distribution is certainly very easy but involves stringent assumptions on the marginal distribution as well as on other distributional aspects.

4 Application

We have selected 10 MSCI world sector indices, choosing daily log-returns (2003/01/01-2015/10/30), for a total of 3347 observations.

The significative lag order d is 1, obtained applying the Kendall's τ autocorrelation coefficients between $r_{i,t}$ and $r_{j,t-d}$ (for $i, j = 1, \dots, 10$, and $d = 1, 2$).

After estimating Kendall's τ and Gaussian VAR (1) coefficients, we have selected 4 of 10 indices characterized by non-linear dependence, in particular: ACWI Industries, ACWI Consumer Staple, ACWI Health Care and ACWI Utilities.

Then, we have compared in terms of AIC and tail dependence measures:

- the Gaussian VAR(1), as benchmark model
- the three-stage Copula-VAR(1), with flexible specifications for the innovations and elliptical copulas for their dependence structure.

In the first stage we have selected t -margins for all MSCI indices (see Table 2), comparing them with normal distributed margins in terms of Akaike's information criterion (AIC). In the second stage of the procedure, given t -margins, t -copulas have been employed in order to model the serial dependence (Table 3), choosing among normal and t -copulas.

We have selected only elliptical copulas in order to show the performance of the model, focusing the attention on the main two families.

Denoting with 1 the ACWI Industries $_t|I_{t-1}$, with 2 the ACWI Consumer Staple $_t|I_{t-1}$, with 3 the ACWI Health Care $_t|I_{t-1}$ and with 4 the ACWI Utilities $_t|I_{t-1}$, the $Copula_i$ (for $i=1,2,3,4$) of Table 3 is a five-dimensional copula for each margin at time t and all margins at $t - 1$.

Finally, given t -copulas of the previous stage, in the third stage we have selected, the last t -copula, with respect to Gaussian copula, to model the contemporaneous dependence (see Table 4) between variables. In this stage, we have estimated a copula among time conditioning variables at same time t . The $Copula_{1234}$ of Table 4, is a four-dimensional t -copula among margins (for each MSCI index) at time t conditioning on all margins at $t - 1$. t -Copula is preferred at each step of the procedure given t -margins.

Indeed, we compare, in terms of AIC values, the Gaussian VAR(1) model (-101057) with the three-stage copula-based VAR model (-104067). The result

shows the best performance of the three-stage copula VAR model over the relevant benchmark model. In Tables 5 and 6 we report the Gaussian VAR parameter results estimated with OLS method.

In Table 7, we report the parameter estimate results of the three stage model, at the second stage. For instance, ρ_1 of the *Copula*₁ is correlation coefficient between the margins corresponding to ACWI Industries at time t and itself at $t-1$, ρ_2 of the *Copula*₁ is the correlation between the margins related to ACWI Industries at t and that of Consumer Staple at $t-1$; in the same way, ρ_1 of the *Copula*₂ is correlation coefficient between the margins of Consumer Staple at t and the margin of ACWI Industries at $t-1$.

Table 8 shows parameter estimates of the copula at the third stage, which are all significant. We can note that the correlation coefficients of the copulas involved in the second stage are almost all significant, contrary to the Gaussian VAR model for which only some autoregressive coefficients are significant (see Table 5). Industries and Health Care sectors, for example, do not show dependence with respect to the past values assumed by the other sectors but only on their own lagged values (previous day). Utilities sector does not display any link with the other sectors. Finally, for the Consumer Staple sector we have found evidence of correlation with its own lag and with one-lagged Utilities. On the other hand, using a three stage approach by means of copulas we can verify a significant time dependence among sectors, with respect to lagged values both own and of the others. The last t -Copula shows tail dependence coefficients between each margin pair (of MSCI index) at time t conditioned on all margins

at $t - 1$ different from zero (as showed in the Gaussian case),

$$\hat{\lambda}_{12} = 0.4135$$

$$\hat{\lambda}_{13} = 0.3471$$

$$\hat{\lambda}_{14} = 0.3390$$

$$\hat{\lambda}_{23} = 0.3980$$

$$\hat{\lambda}_{24} = 0.3696$$

$$\hat{\lambda}_{34} = 0.2853$$

Also in this case, we observe a strong, both positive and negative, dependence between sectors in the tails of the joint distribution. Extreme negative events (positive) that occur in a sector a day before can have a great negative (positive) impact on the other sector values the day after.

5 Concluding remarks

Financial returns do not show any Gaussian behavior, both in univariate and multivariate cases. Auto-correlation and cross-covariance are also documented in many papers and, in addition, a non-linear cross-dependence emerges over the time. Besides, the dependence in the tails of joint distributions of financial variables is an aspect that we have not to underestimate.

We have provided a simple method, supported by simulated data, to create models for multiple time series which involve both non-linear temporal and cross-dependence dynamics, considering also a measure of tail co-movements. To this end, we have made use of copulas, extending the three-stage pseudo

maximum likelihood estimation in order to obtain a non-linear general VAR model in high dimensions, in an easier way than using a unique loglikelihood function, giving also a contribution on the recent literature.

To prove the performance of the model on real data, we have investigated the dynamics of the serial and cross-sectional dependence among MSCI world sector indices with two different methodologies in a linear and non-linear framework, comparing the traditional Gaussian Vector AutoRegressive model (VAR) with a more flexible general VAR model. We conclude that the three-stage approach is the best solution to create non-linear model to analyse multiple time series avoiding the problems related to rigid methods typical of benchmark linear models.

References

- Abegaz, F., Naik-Nimbalkar, U. V., 2007. Modeling statistical dependence of Markov chains via copula models. *Journal of Statistical Planning and Inference* 04, 028.
- Bianchi C., Carta A., Fantazzini D. De Giuli M.E., Maggi M.A., 2010. A Copula-VAR-X approach for Industrial Production, *Applied Economics* 42 (25), 3267–3277.
- Brechmann E.C., Czado C., 2015. COPAR multivariate Time Series Modeling Using the Copula Autoregressive Model, *Applied Stochastic Models in Business and Industry* 31 (4), 495–514.

- Chen, X., Fan, Y., 2006a. Estimation of copula-based semiparametric time series models. *Journal of Econometrics* 130, 307-335.
- Cherubini U., Luciano E., Vecchiato W., 2004. *Copula Methods in Finance*, John Wiley & Sons, England.
- Demarta S., McNeil A.J., 2005. The t copula and related copulas. *International Statistical Review*.
- DeMiguel V., Nogales F.J., Uppal R., 2014. Stock Return Serial Dependence and Out-of-Sample Portfolio Performance, *Review of Financial Studies* 27 (4), 1031–1073.
- De Luca, G., Riviello G., 2012. Multivariate tail dependence coefficients for Archimedean Copulae. A. Di Ciaccio et al. (eds.), *Advanced Statistical Methods for the Analysis of Large Data-Sets, Studies in Theoretical and Applied Statistics*, Springer.
- Dobrić, J., Schmid, F., 2005: Non parametric estimation of the lower tail dependence λ_L in bivariate copulas. *Journal of Applied Statistics* **32**, 387–407
- Fama E.F., French K.R., 1998. Value versus Growth: The International Evidence, *The Journal of Finance* 53 (6), 1975–1999 .
- Hamilton J.D. 1994. *Time Series Analysis*, Princeton University Press.
- Joe H., Xu J.J., 1996. The estimation method of inference functions for margins for multivariate models, Technical Report, 166, Department of Statistics, University of British Columbia.

- Joe H., 1997. *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability. London: Chapman and Hall.
- Kotz S., Nadarajah S., 2004. *Multivariate t Distributions and Their Applications*, Cambridge University Press.
- Lo A.W., MacKinlay A.C., 1990. When Are Contrarian Profits Due to Stock Market Overreaction? *Review of Financial Studies* 3 (2), 175–205, Springer.
- Longin F., Solnik B., 2001. Extreme correlation of international equitymarkets. *Journal of Finance* 56, 649-676.
- Nelsen R.B., 2006. *An Introduction to Copulas*, Springer.
- Pickands J., 1981: Multivariate extreme value distributions. *Proceedings of the 43rd Session ISI (Buoneos Aires)*,399-409.
- Yi W., Liao S.S., 2010. Statistical properties of parametric estimators for Markov chain vectors based on copula models. *Journal of Statistical Planning and Inference*, 140, 1465-1480.

Table 1: True (λ) and average estimated ($\bar{\lambda}$) tail dependence coefficients (standard errors in brackets)

Pair	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
λ	0.0859	0.0835	0.0924	0.0707	0.0938	0.0859
$\bar{\lambda}$	0.0499	0.0479	0.0548	0.0388	0.0559	0.0498
	(0.0132)	(0.0127)	(0.0137)	(0.0111)	(0.0143)	(0.0131)

Table 2: AIC comparison for margins

	Normal		Student's t	
	LogL	AIC	LogL	AIC
MSCI ACWI				
Industries	10411	-20818	10811	-21616
Consumer Staple	11621	-23237	12022	-24037
Health Care	11169	-22334	11503	-23001
Utilities	10986	-21968	11438	-22869

Table 3: AIC comparison for serial dependence copulas

	Normal		Student's t	
	LogL	AIC	LogL	AIC
$Copula_1$	5476	-10932	6081	-12140
$Copula_2$	5417	-10815	5995	-11968
$Copula_3$	5418	-10816	6004	-11986
$Copula_4$	5421	-10821	6000	-11978
$Copula_{t-1}$	5410	-10807	5900	-11786

Table 4: AIC comparison for the contemporaneous dependence copula

	Normal		Student's t	
	LogL	AIC	LogL	AIC
$Copula_{1234}$	5484	-10956	5823	-11632

Table 5: Gaussian VAR: OLS parameter estimates

Explanatory Variables $_{t-1}$	Dependent variable $_t$			
	ACWI Industries	ACWI Consumer Staple	ACWI Health Care	ACWI Utilities
	Estimate (st.error)	Estimate (st.error)	Estimate (st.error)	Estimate (st.error)
ACWI Industries	0.174* (0.033)	-0.004 (0.023)	-0.031 (0.026)	-0.018 (0.028)
ACWI Consumer Staple	0.054 (0.057)	0.0876* (0.041)	-0.019 (0.046)	0.059 (0.050)
ACWI Health Care	0.016 (0.040)	0.037 (0.028)	0.155*** (0.032)	0.051 (0.034)
ACWI Utilities	-0.0512437 (0.038)	-0.058* (0.027)	-0.054 (0.031)	0.017 (0.033)

Signif.codes : 0***0.001**0.01*0.05'

Table 6: Gaussian VAR: correlation matrix of residuals

Variables	Correlations			
	ACWI Industries	ACWI Consumer Staple	ACWI Health Care	ACWI Utilities
ACWI Industries	1	0.8370	0.7713	0.7911
ACWI Consumer Staple	0.8370	1	0.8373	0.8293
ACWI Health Care	0.7713	0.8373	1	0.7538
ACWI Utilities	0.7911	0.8293	0.7538	1

Table 7: Three stage model: parameter estimates of serial dependence copulas

Parameter	<i>Copula</i> ₁		<i>Copula</i> ₂		<i>Copula</i> ₃		<i>Copula</i> ₄	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
ρ_1	0.185***	0.018	0.047*	0.018	0.016	0.018	0.039*	0.018
ρ_2	0.149***	0.018	0.057**	0.018	0.024	0.018	0.049**	0.018
ρ_3	0.139***	0.018	0.048**	0.018	0.054**	0.018	0.046*	0.018
ρ_4	0.149***	0.018	0.048**	0.018	0.026	0.018	0.069***	0.018
ρ_5	0.814***	0.005	0.814***	0.005	0.814***	0.005	0.814***	0.005
ρ_6	0.750***	0.007	0.751***	0.007	0.752***	0.007	0.750***	0.007
ρ_7	0.750***	0.007	0.751***	0.007	0.750***	0.007	0.750***	0.007
ρ_8	0.805***	0.006	0.806***	0.006	0.808***	0.006	0.805***	0.006
ρ_9	0.783***	0.006	0.784***	0.006	0.789***	0.006	0.783***	0.006
ρ_{10}	0.697***	0.008	0.699***	0.008	0.700***	0.008	0.697***	0.008
<i>df</i>	4.215***	0.168	4.343***	0.177	4.326***	0.174	4.308***	0.175

Signif.codes : 0***0.001**0.01*0.05'

*Copula*_{*i*} (*i* = 1, 2, 3, 4) is a five-dimensional *t*-copula among a margin (for each MSCI index) at *t* and all margins at *t* - 1

Table 8: Three stage model: parameter estimates of contemporaneous dependence copula

<i>Copula</i> ₁₂₃₄		
Parameter	Estimate	Std. Error
ρ_{12}	0.812***	0.005
ρ_{13}	0.756***	0.007
ρ_{14}	0.748***	0.007
ρ_{23}	0.800***	0.006
ρ_{24}	0.777***	0.006
ρ_{34}	0.691***	0.008
<i>df</i>	6.274***	0.415

Signif.codes : 0***0.001**0.01*0.05'

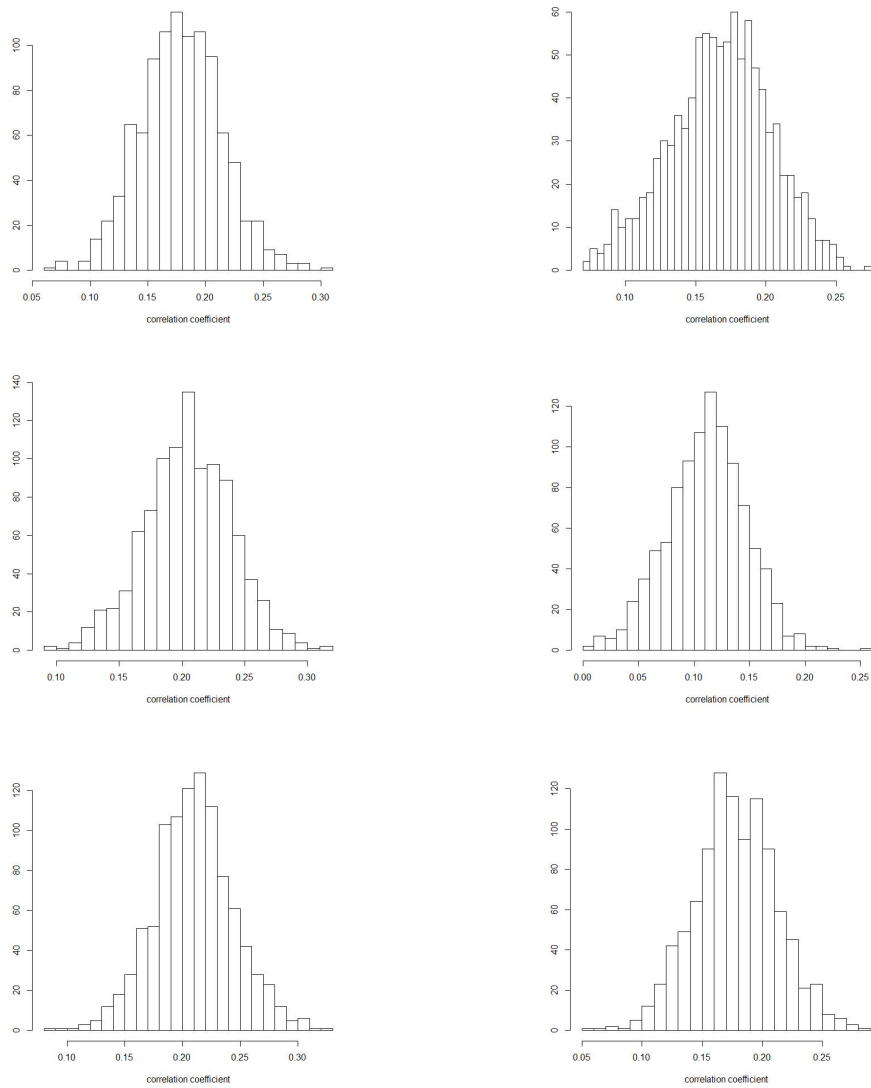


Figure 1: Histograms of correlation coefficients of the contemporaneous copula

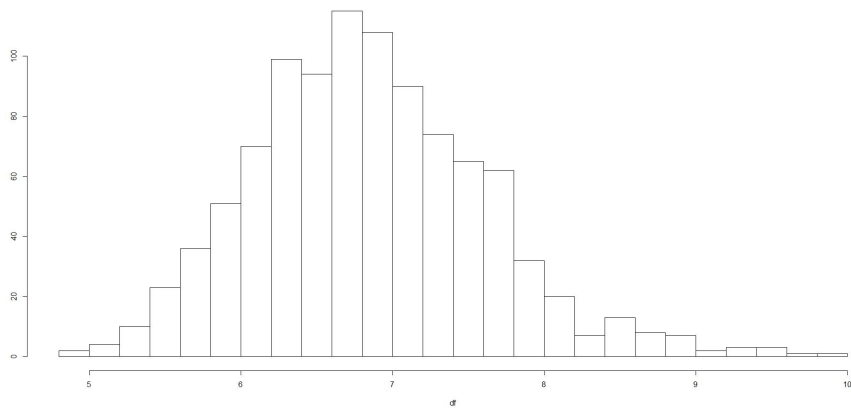


Figure 2: Histogram of degrees of freedom of the contemporaneous copula

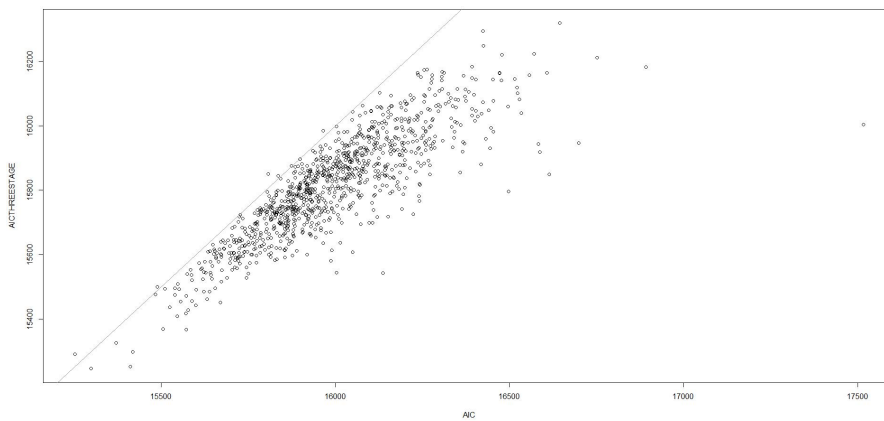


Figure 3: AIC comparison between the Gaussian VAR model (average=15993) and the three-stage copula-based VAR approach (average=15817)

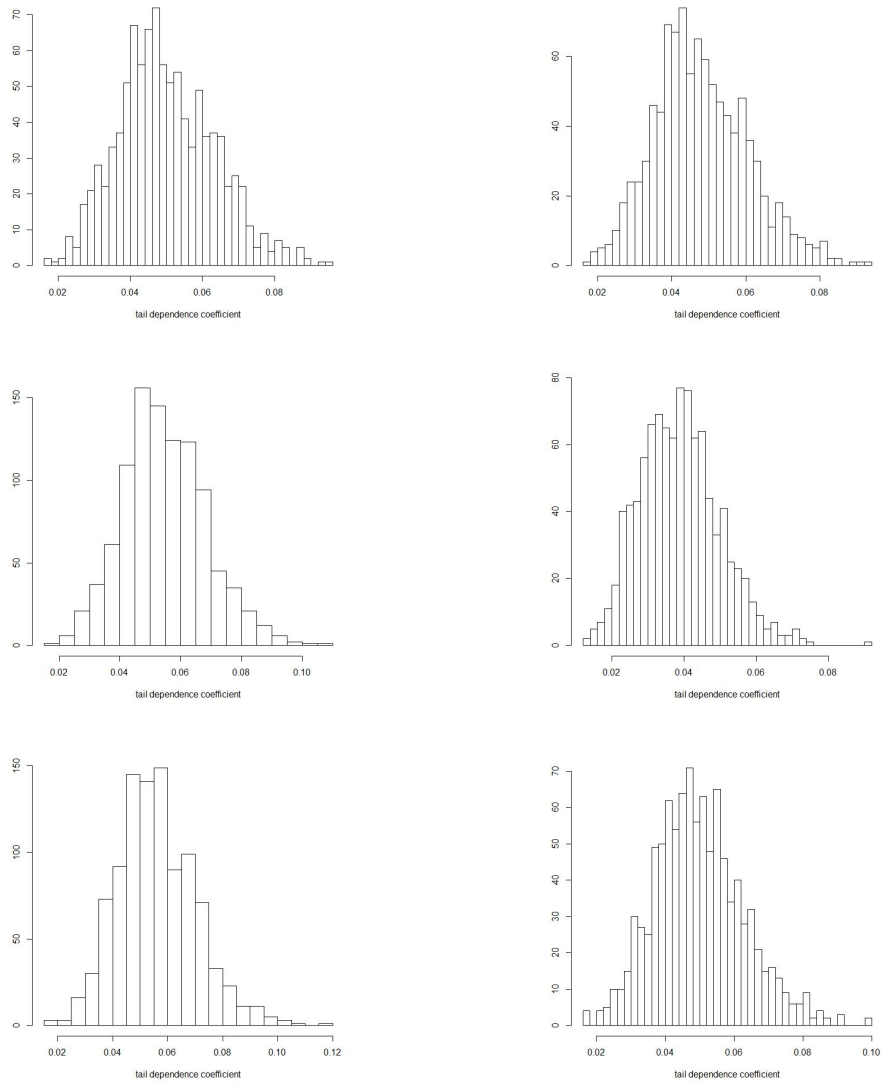


Figure 4: Histogram of tail dependence coefficients of the contemporaneous copula