Clustering of financial time series in extreme scenarios

Fabrizio Durante and Roberta Pappadà

Abstract A methodology is presented for clustering financial time series in extreme scenario. The procedure is based on the calculation of some suitable pairwise conditional Spearman's correlation coefficients. It does not assume any parametric model describing the time series under investigation, but only relies on the assumption that they follows a multivariate copula-GARCH model.

Key words: Cluster Analysis, Copula, Spearman's correlation, Tail dependence.

1 Introduction

In portfolio risk analysis a current practice for minimizing the whole risk consists of adopting some *diversification techniques* that are based, roughly speaking, on the selection of different assets from markets and/or regions that one believes to be weakly (or negatively) correlated. Such an approach tries to reduce the impact of joint losses that might occur simultaneously in different markets.. To this end, suitable cluster techniques for multivariate time series have been proposed in the literature in order to give a guideline to practitioners for the selection of a suitable portfolio. Such techniques span from the use of correlation coefficient (see, for instance, [1]) to the use of techniques based on the comparisons among the coefficients of the underlying processes (see, for instance, [5, 11, 12]).

However, it has been stressed several times that diversification principle may fail when there is some *contagion* among the markets under consideration (see [2, 4, 7]),

Fabrizio Durante
School of Economics and Management, Free University of Bozen–Bolzano, Bolzano (Italy)
e-mail: fabrizio.durante@unibz.it

Roberta Pappadà
Department of Statistical Sciences, University of Padua, Padua (Italy)
e-mail: pappada@stat.unipd.it

namely when the positive association among the markets increases in crisis period with respect to tranquil periods. To this end, it could be useful to introduce some clustering methods that focus their attention to the behaviour of financial markets in presence of risky scenarios. An innovative work in this direction has been recently done by De Luca and Zuccolotto [3], who proposed a clustering procedure that aims at grouping time series with an association between extremely low values, measured by a tail dependence coefficient.

Starting with the ideas of [3], this note aims at presenting some new clustering procedures for extreme scenario. Such a methodology is grounded on the conditional (Spearman's) correlation coefficient between time series. It aims at creating cluster of time series that are homogeneous, in the sense that they tend to be comonotone in their extreme low values (where the degree of extremeness is specified by a given threshold α).

2 The methodology

Let $(x_t^i)_t$ be a vector of financial time series $(i = 1, 2, \dots, d)$ representing the returns of different assets and/or stock indices. In order to provide some cluster procedure we are interested in determining a suitable distance or dissimilarity measure among them.

To this end, we make use of the Spearman's correlation coefficient ρ_S. Compared with standard linear correlation, we recall that ρ_S is invariant under rank transformation of the marginals, and, hence, it is more adapt to describe non-linear behaviour of time series (see, for instance, [6]).

Moreover, since we are interested in the tail of the time series, we will restrict to consider ρ_S conditional on extreme observations, i.e. we calculate ρ_S conditional on the fact that the time series are below a given threshold α defined in a suitable way.

As said, e.g., in [7], the calculation of ρ_S for conditional events can be biased by the volatility effects of the individual time series, regardless any other influence of the dependence among them.

To remove such an effect, we need to some additional assumptions about the stochastic model from which the time series are generated.

Starting with a copula approach [9], we suppose that the joint model of $(x_t^1, \dots, x_t^d)_t$ may be determined in the following way: the marginal times series are supposed to be derived from an AR(1)–GARCH(1,1) model with innovation distribution being symmetric Student distribution; while the residuals $(x_t^1, \dots, x_t^d)_t$ are independent and identically distributed with dependence structure given by an unknown (but fixed) copula C.

Basically, such a procedure applies a filter to the individual time series in order to remove heteroscedastic effects (see, for instance, [10]).

Now, the analysis of the extremal correlation among the time series relies on the calculation of the pairwise Spearman's correlation among residuals in a suitable tail region of the copula C (or better, of its empirical version). In particular, since ρ_S is scale-invariant, we may adopt the following procedure:

- Rescale the residuals from each time series to the interval $[0,1]$ (by using, for instance, their empirical cumulative distribution function) by obtaining $(z_t^1,\dots,z_t^d)_t$.
- Fix a given level α (usually, $\alpha = 0.05, 0.10, 0.25$), which is a threshold denoting the "degree" of risk of the scenario we are considering.
- For every $i \neq j$ calculate the conditional ρ_S of $(z_t^i,,z_t^j)_t$ given the fact that $z_t^i \leq \alpha$ and $z_t^j \leq \alpha$. Let us call this value $\rho_S^{ij}(\alpha)$.

Now, for $i,j = 1,\dots,d$, we may define the matrix $\left(\sqrt{2(1-\rho_S^{ij}(\alpha)}\right)_{i,j}$, which forms a distance matrix according to [1].

The matrix above defined induces clusters amongst the d time series of financial returns by means of a suitable procedure. For our purposes, we consider the hierarchical agglomerative clustering technique frequently used in practice. The idea is to cluster time series in homogeneous groups according to the distance matrix adopted, in the sense that the elements within the resulting clusters are expected to show a similar behaviour in extreme scenario.

As known, hierarchical agglomerative algorithms start from the finest partition possible (each observation forms a cluster) and each level merges a selected pair of clusters into a new cluster according to the definition of the distance between two groups. This sequence of nested partitions is best visualized as a top-down tree called a dendrogram, such that the dissimilarity between merged clusters is monotone increasing with the level of the merger.

Among all the agglomerative strategies we consider the four most common clustering procedures which differ in the computation of the distance between two groups: single linkage, complete linkage, average linkage and Ward's method (see, for instance,[8]). In particular, the single linkage algorithm defines the distance between two groups as the smallest value of the individual distances; the complete linkage algorithm merges together groups by considering the largest (individual) distances; the average linkage (weighted or unweighted) represents a compromise between the two preceding algorithms, since it computes an average distance. The Ward clustering method differs from the first three in the unification procedure: it aims to unify groups such that the variation inside these groups does not increase too drastically. The resulting groups are as homogeneous as possible. Among all the algorithms described, the Ward method is expected to give the best results.

In the talk, we will present a small simulation study about the behaviour of the model when the dependence among the time series is described via some known asymmetric copula models. Moreover, the described procedure will be applied to European stock markets by showing its practical economic consequences.

3 Conclusions

We have presented a methodology that can be applied in order to obtain clusters of financial time series that take into account the dependence in risky scenario. The procedure is non-parametric, but it assume the existence of a suitable GARCH–copula model that could describe the joint behaviour of the time series. According to [3], we expect that the procedure could be applied in portfolio analysis, with particular emphasis on portfolio selection.

Acknowledgements The first author acknowledges the support of Free University of Bozen-Bolzano, School of Economics and Management, via the project "Risk and Dependence".

References

1. Bonanno, G., Caldarelli, G., Lillo, F., Miccichè, S., Vandewalle, N., Mantegna, R.: Networks of equities in financial markets. Eur. Phys. J. B **38**(2), 363–371 (2004)
2. Bradley, B., Taqqu, M.: Framework for analyzing spatial contagion between financial markets. Finance Letters **2**(6), 8–16 (2004)
3. De Luca, G., Zuccolotto, P.: A tail dependence-based dissimilarity measure for financial time series clustering. Adv. Data Analysis and Classification **5**(4), 323–340 (2011)
4. Durante, F., Jaworski, P.: Spatial contagion between financial markets: a copula-based approach. Appl. Stoch. Models Bus. Ind. **26**(5), 551–564 (2010)
5. D'Urso, P., Maharaj, E.: Autocorrelation-based fuzzy clustering of time series. Fuzzy Sets and Systems **160**(24), 3565–3589 (2009)
6. Embrechts, P., McNeil, A.J., Straumann, D.: Correlation and dependence in risk management: properties and pitfalls. In: M. Dempster (ed.) Risk management: value at risk and beyond, pp. 176–223. Cambridge University Press, Cambridge (2002)
7. Forbes, K.J., Rigobon, R.: No contagion, only interdependence: measuring stock market comovements. The Journal of Finance **57**(5), 2223–2261 (2002)
8. Härdle, W., Simar, L.: Applied Multivariate Statistical Analysis. Springer (2012). 3rd ed
9. Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.): Copula Theory and its Applications, *Lecture Notes in Statistics - Proceedings*, vol. 198. Springer, Berlin Heidelberg (2010)
10. Nikoloulopoulos, A., Joe, H., Li, H.: Vine copulas with asymmetric tail dependence and applications to financial return data. Computational Statistics & Data Analysis, in press (2012)
11. Otranto, E.: Clustering heteroskedastic time series by model-based procedures. Computational Statistics & Data Analysis **52**(10), 4685–4698
12. Piccolo, D.: A distance measure for classifying ARIMA models. Journal of Time Series Analysis **11**(2), 153–164 (1990)