Journal of Multivariate Analysis 123 (2014) 96–110

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Measuring association and dependence between random vectors

Oliver Grothe [a,*], Julius Schnieders [a], Johan Segers [b]

[a] University of Cologne, Department of Economic and Social Statistics, Albertus-Magnus-Platz, 50923 Cologne, Germany

[b] Université catholique de Louvain, Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

ARTICLE INFO

Article history:
Received 25 January 2013
Available online 9 September 2013

AMS subject classifications:
62H20
62G30

Keywords:
Association
Copula
Dependence
Kendall's tau
Spearman's rho
U-statistic

ABSTRACT

Measures of association are suggested between two random vectors. The measures are copula-based and therefore invariant with respect to the univariate marginal distributions. The measures are able to capture positive as well as negative association. In case the random vectors are just random variables, the measures reduce to Kendall's tau or Spearman's rho. Nonparametric estimators, based on ranks, for the measures are derived. Their large-sample asymptotics are derived and their small-sample behavior is investigated by simulation. The measures are applied to characterize strength and direction of association of northern and southern European bond markets during the recent Euro crisis as well as association of stock markets with bond markets.

1. Introduction

In many applications, measuring the association between different random vectors is of special interest. However, most of the literature on association of random vectors concerns association *within* a random vector, that is, among its univariate components. Classical measures for bivariate vectors are Spearman's rho, Kendall's tau, Blomqvist's beta, Gini's gamma and several lesser known measures. For most of these, multivariate extensions to higher dimensions exist [22,24,32]; see [33] for a survey. A measure specifically designed for the multivariate case is Kendall's W coefficient of concordance [25], recently studied using copulas by Grothe and Schmid [13] and Marozzi [26].

The question of measuring dependence and association *between* random vectors is less investigated. It starts with the canonical correlation measure of Hotelling [20], a measure of linear dependence between random vectors. Another classical measure is the RV coefficient [10,31]. Most articles concern tests of independence between random vectors rather than actually measuring the association. For this purpose, matrices of Spearman and Kendall correlation coefficients [9] or other rank-based statistics [28] or averages of Spearman's rho and Kendall's tau [17] are used. Other statistics are based on distances of empirical copulas [29] or densities [36,35].

Instead of testing for independence of random vectors, the focus of this paper is the development of margin-free measurement of association *between* random vectors. Such measures may be informative when analyzing comovements or contagion of financial markets [12]. In such applications, the random vectors correspond to, e.g., groups of assets from different markets (like stock markets, commodity and energy markets) or of economic factors as unemployment rate, GDP or inflation

* Corresponding author.
E-mail addresses: grothe@statistik.uni-koeln.de (O. Grothe), schnieders@wiso.uni-koeln.de (J. Schnieders), johan.segers@uclouvain.be (J. Segers).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.08.019

rate. Being invariant with respect to increasing transformations of the margins, the measures then quantify the association between groups of variables as a whole instead of the association between individual assets or factors.

The dependence between markets may be of positive or of negative nature and it will be important to be able to make this distinction. Consider for example a vector of returns of assets from the banking sector and a vector of returns from the automobile industry. Association between these vectors will most likely be positive. On the other hand, consider a vector of interest rates and a vector of assets returns. Here, standard economic theory predicts a negative association, since, usually, interest rates are assumed to increase when the stock market is bearish (prices trending down) and decrease when the stock market is bullish (prices trending up). Measures which are interesting for our purpose should therefore be able to distinguish between positive and negative association. As canonical correlation, the RV coefficient and distance correlation are always positive, they are of limited use for our purpose.

We therefore develop two measures of association, generalizations of Spearman's rho and Kendall's tau, which both have the desired properties. In case the univariate marginal distributions of the random vectors are continuous, which is the focus of this paper, the measures depend only on the underlying copulas and are therefore margin-free. The basic idea of the measures is that independence of vectors of length p and q means that the joint $(p+q)$-variate cumulative distribution function (cdf) factorizes into the product of the p-variate and q-variate marginal cdfs. In general, the difference between the $(p+q)$-variate cdf and the product of p-variate and q-variate marginal cdfs contains information on the strength of the dependence and the nature of the association, positive or negative. Integrating out this difference and normalizing it appropriately yields our measures of association.

The definitions of our association measures between random vectors are inspired by the ones for random variables, notably Spearman's rho and Kendall's tau. Association is measured by a normalized integral of the difference of the joint distribution function of the two random vectors and their distribution function under independence, i.e. the product of their marginal distribution functions. In [33], multivariate versions of Spearman's rho for association within a random vector are defined via such integrals as well.

The existing and new measures of association somehow average out the pairwise associations of the variables involved. This means that if the measures are positive, positive association is predominant within the variables and vice versa. However, positive values of the measures do not mean that all pairwise associations involved are positive. Neither do negative values of the measures imply negativity of all pairwise associations.

A crucial point when measuring association in this way is how exactly to integrate out the difference between the joint and the marginal cdfs. We propose weightings which lead to generalizations of Spearman's rho and Kendall's tau. Furthermore, the weightings ensure desirable properties with respect to the concordance ordering, properties which will be discussed in detail. Concordant variables tend to be all large together or all small together [22]. For fixed p-variate and q-variate marginal cdfs, more concordant joint $(p+q)$-variate distributions lead to larger values of our measure. Furthermore, if the joint $(p+q)$-variate copula is elliptical, the association measures are increasing functions of the pairwise correlations between components of the first and the second random vectors.

We propose nonparametric inference procedures for our association measures based upon U-statistics. The use of U-statistics in this context is quite natural; for Spearman's rho, it goes back to Hoeffding [19, p. 318]. Because the measures are copula based, the inference procedures depend on the data only through the ranks, lending some robustness to the methodology. We provide explicit estimators, establish their asymptotic normality, and show how to estimate their asymptotic variance in a computationally efficient way. The small-sample performance is assessed through Monte Carlo simulations and the methodology is illustrated via a case study measuring dependence between bond and stock markets as well as between bond markets of northern and southern European countries.

The structure of the paper is as follows: In Section 2, we describe the notation and terminology used within the paper. In Section 3, we define and motivate population versions of our measures. Estimators for the measures and their variances are introduced in Section 4. Section 5 contains the simulation study and an empirical examples with financial data. Software for the estimators, written in R and C, is available on request. Section 6 concludes.

2. Preliminaries

To motivate our approach, we recall Spearman's rho and Kendall's tau for random variables X and Y with continuous distributions. Let their joint and marginal distribution functions be denoted by $F(x) = \Pr[X \le x]$, $G(y) = \Pr[Y \le y]$ and $H(x, y) = \Pr[X \le x,\ Y \le y]$, for $x, y \in \mathbb{R}$. Then Spearman's rho can be expressed as

$$\rho_S = 12 \int_{\mathbb{R}} \int_{\mathbb{R}} \{H(x, y) - F(x)\, G(y)\}\, \mathrm{d}x\, \mathrm{d}y$$

while Kendall's tau is given by

$$\tau = 4 \left(\int_{\mathbb{R}^2} H(x, y)\, \mathrm{d}H(x, y) - \int_{\mathbb{R}} \int_{\mathbb{R}} F(x)\, G(y)\, \mathrm{d}F(x)\, \mathrm{d}G(y) \right).$$

For both measures, there exist generalizations to d-dimensional random vectors [33]. In all cases, the measures of association are constructed by averaging out the difference between the joint distribution function, $H(x, y)$, and the distribution function in case of independence, $F(x)\, G(y)$.