Physica A 390 (2011) 4251–4259

Contents lists available at SciVerse ScienceDirect

Physica A

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

A copula approach on the dynamics of statistical dependencies in the US stock market

Michael C. Münnix [a,b,*], Rudi Schäfer [a]

[a] Fakultät für Physik, Universität Duisburg-Essen, 47048 Duisburg, Germany

[b] Center of Polymer Studies, Department of Physics, Boston University, USA

ARTICLE INFO

Article history:
Received 10 January 2011
Received in revised form 5 April 2011
Available online 29 June 2011

Keywords:
Financial correlations
Statistical dependencies
Copula
Market dynamics

ABSTRACT

We analyze the statistical dependence structure of the S&P 500 constituents in the 4-year period from 2007 to 2010 using intraday data from the New York Stock Exchange's TAQ database. Instead of using a given parametric copula with a predetermined shape, we study the empirical pairwise copula directly. We find that the shape of this copula resembles the Gaussian copula to some degree, but exhibits a stronger tail dependence, for both correlated and anti-correlated extreme events. By comparing the tail dependence dynamically to the market's average correlation level as a commonly used quantity we disclose the average level of error of the Gaussian copula, which is implied in the calculation of many correlation coefficients.

1. Introduction

The measurement of statistical dependence is often broken down to the calculation of a correlation coefficient, such as the Pearson coefficient [1] or the Spearman coefficient [2]. Correlation coefficients are widely used in various disciplines of science. It is also often included in financial modeling, e.g., in the Capital Assets Pricing Model (CAPM) [3] or Noh's model [4].

The Pearson correlation coefficient, however, only accounts for linear statistical dependence assuming that the observables are nearly normal distributed. Due to the central limit theorem, this might be justified in some cases, but often the statistical dependence is much more complex. In these cases, the statistical dependence cannot be represented by a single number. The joint probability distribution, of course, holds all information of the statistical dependence. Certainly, the joint probability distribution also contains the individual marginal probability distributions. These can have different shapes depending on the underlying process. The statistical dependence of different systems usually cannot be directly compared with this approach.

Copulae, first introduced by Sklar in 1959 [5,6], permit a separation between the pure statistical dependence and the marginal probability distributions. This allows to compare the statistical dependence of diverse systems.

The usage of copulae is well established in statistics and finance. There are many classes of analytical copula functions that meet various properties [7]. Several studies of financial markets are devoted to developing suitable copulae or fitting existing ones to empirical data [8–10] or are based on a small subset of assets [11].

In this study, we choose a different approach. We perform a large-scale empirical study to disclose the structure of the average pairwise copula of the US stock returns. As the copula does not depend on the shape of the return distribution, we are able to average over the copula of different stock pairs although their marginal distributions' shape may differ, i.e., exhibits stronger or weaker tails. In particular, we study the intraday stock market returns of the 428 continuous S&P 500

* Corresponding author at: Fakultät für Physik, Universität Duisburg-Essen, 47048 Duisburg, Germany. Tel.: +49 203 379 4727; fax: +49 203 379 4732.
E-mail address: michael@muennix.com (M.C. Münnix).

0378-4371/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.06.032

constituents in 2007–2010 based on intraday data from the New York Stock Exchange's TAQ database. This empirical study enables us to disclose the average level of error involved in the Gaussian copula. We then compare the tail dependence of the empirical copula and the Gaussian copula dynamically based on a moving window of 2-weeks. We map this tail dependence to the market's average correlation level and thereby provide an error estimate for the Gaussian copula based on the current general market situation.

This article is organized as follows. In Section 2, we give a brief introduction to the concept of copulae. We discuss the average pairwise copula in Section 3, which is extended by dynamical aspects in Section 4. We conclude our results in Section 5.

2. Copulae

The basic concept is simple: let a and b be two random variables with probability densities $f_a(x)$ and $f_b(x)$ and cumulative distributions $F_a(x)$ and $F_b(x)$, with

$$\int_{-\infty}^{+\infty} f_a(x) = 1, \tag{1}$$

$$F_a(x) = \int_{-\infty}^{x} f_a(x')\,\mathrm{d}x', \tag{2}$$

and analogously for b. Further, let $f_{a,b}(x, y)$ be the joint probability density and $F_{a,b}(x, y)$ be the joint cumulative distribution. The inverse cumulative distribution function F^{-1} is called quantile function. For example, 5% of all random samples are smaller or equal than $F_a^{-1}(0.05)$ represents the value. This evidently gives,

$$F_a\left(F_a^{-1}(\alpha)\right) = \alpha. \tag{3}$$

$F^{-1}(\alpha)$ is also called the α-quantile. The copula $\mathrm{Cop}_{a,b}(u, v)$ is defined as the cumulative joint distribution of quantiles,

$$\mathrm{Cop}_{a,b}(u, v) = F_{a,b}\left(F_a^{-1}(u), F_b^{-1}(v)\right). \tag{4}$$

The copula density $\mathrm{cop}_{a,b}(u, v)$ is consequently defined by

$$\mathrm{cop}_{a,b}(u, v) = \frac{\partial^2}{\partial u \partial v}\mathrm{Cop}_{a,b}(u, v). \tag{5}$$

As the quantile functions F^{-1} are scale free, the copula does not depend on the underlying marginal distributions. It only contains the pure statistical dependence. Thus, by obtaining the appropriate copula of a system, one can simply interchange the marginal distributions without any changes in the copula. This is very useful if the marginal distributions change for some reason, but the statistical dependence remains the same. We can rebuild the joint cumulative distribution from the copula and the individual distributions by

$$F_{a,b}(x, y) = \mathrm{Cop}_{a,b}\left(F_a(x), F_b(y)\right). \tag{6}$$

3. Average copula

To calculate the cumulative copula from empirical data of two return time series r_1 and r_2, we use

$$\mathrm{Cop}_{r_1,r_1}(u, v) = \frac{1}{T}\sum_{t=1}^{T} 1_U(r_1(t)) \times 1_V(r_2(t)), \tag{7}$$

where T is the length of the time series. 1_U and 1_V are indicator functions relating to the sets

$$U = \left\{x \mid x \le F_1^{-1}(u)\right\}, \tag{8}$$

$$V = \left\{y \mid y \le F_2^{-1}(v)\right\}. \tag{9}$$

The quantile function F^{-1} on empirical data is given by

$$F_1^{-1}(u) = \begin{cases} \inf\{x \mid F_1(x) \ge u\} & 0 < u \le 1 \\ \sup\{x \mid F_1(x) = u\} & u = 0, \end{cases} \tag{10}$$

and analogously for r_2. We define $F_1(x)$ empirically as the percentage of the portion that is smaller or equal to x compared to the total amount of values. When calculating the empirical copula density, it is useful to first define a resolution of the 2D grid, e.g. $m = 50$. On this $m \times m$ grid, we can calculate the copula by

$$\mathrm{cop}_{r_1,r_1}\left(\frac{i}{m}, \frac{j}{m}\right) = \frac{1}{T}\sum_{t=1}^{T} 1_{\bar{U}_i}(r_1(t)) \times 1_{\bar{V}_j}(r_2(t)) \quad i, j \in 1 \ldots m \tag{11}$$