Physica A 463 (2016) 246–261

Contents lists available at ScienceDirect

Physica A

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

Comparing minimum spanning trees of the Italian stock market using returns and volumes

Paolo Coletti

Faculty of Economics, Free University of Bozen Bolzano, piazza dell'Università 1, 39100 Bolzano, Italy

HIGHLIGHTS

- An extended network analysis of 100 Italian stocks based on returns.
- Special care is taken regarding data sources, in particular corrections for dividends and capital operations.
- Four methods are adopted using symbolization rules and volumes.
- Reliability and centrality measures are used to compare methods and industry sectors.
- Petroleum and utilities sectors form clusters, insurances stocks are central.

ARTICLE INFO

Article history:
Received 9 March 2016
Received in revised form 30 May 2016
Available online 25 July 2016

Keywords:
Minimum spanning tree
Italian stock market
Correlation network
Symbolic time-series analysis

ABSTRACT

We have built the network of the top 100 Italian quoted companies in the decade 2001–2011 using four different methods, comparing the resulting minimum spanning trees for methods and industry sectors. Our starting method is based on Person's correlation of log-returns used by several other authors in the last decade. The second one is based on the correlation of symbolized log-returns, the third of log-returns and traded money and the fourth one uses a combination of log-returns with traded money. We show that some sectors correspond to the network's clusters while others are scattered, in particular the trading and apparel sectors. We analyze the different graph's measures for the four methods showing that the introduction of volumes induces larger distances and more homogeneous trees without big clusters.

1. Introduction

One of the main strategies when building equity portfolios in finance is diversification [1], investing in stocks which are expected to behave very differently, ideally strongly negatively correlated, to minimize the portfolio's variance and thus the risk. The basic method to identify equities with a similar behavior is to arrange companies by industrial sector [2] and invest picking equities from different sectors, a strategy which undoubtedly protects the investor from a crisis which may affect a specific sector. However, there is no guarantee that having a different main activity leads to different behaviors in equity prices. Building a network of companies is becoming a new approach to identifying relations among companies, in particular for price relations [3–5].

The starting idea of building a network based on the log difference of equity prices comes from Mantegna [3], who proposes building a correlation matrix of log-returns, an induced distance and consequently a network of companies. As

E-mail address: paolo.coletti@unibz.it.

http://dx.doi.org/10.1016/j.physa.2016.07.029
0378-4371/

the correlation matrix is dense and the resulting network would have an overwhelming amount of linkages, Mantegna suggests building a minimum spanning tree (MST) [6], which can give an overview of the structure without cycles, which is therefore very comprehensible for professionals.

Since this initial work, several different methods have been proposed and analyzed, comparing their results for different stock markets using measures borrowed from graph theory. There are two main streams: the first one remains with MST, trying to deal with the problem that its displayed information is partial, as it sacrifices for readability purposes some potentially strong relations and at the same time displays very weak edges. Tumminello [7] and Kantar [8] couple the MST with a linkage reliability measure, Brida and Risso [9] build confidence intervals through a Monte Carlo simulation, Coronnello et al. [10] compare results obtained building the MST with Kruskal's algorithm with other similar methods using distance among clusters instead of nodes (called ALCA) and building a variation of MST which includes cliques and loops (called PMFG). A similar strategy is used by Tumminello et al. [11] building a PMGF with the maximum number of linkages to keep it topologically planar. Onnela et al. [12–14] use the MST including linkages above a certain correlation threshold, a very simple variation which recovers hidden strong relations. Majapa and Gossel [15], Peron et al. [16] and Sandoval [17] use time shifting windows to check the persistence of linkages. The other stream turns back to the full network, dealing with the problem of too many linkages. Some authors, such as Huang et al. [18] Molgedey and Ebeling [19], analyze the entire network using graph measures without displaying it. Others, such as Tse et al. [20], Nobi et al. [21] and Heiberger [22], set a threshold and display only a partial network, or Kwapień et al. [23] set a threshold to the correlation matrix to spot clusters before switching to an MST. Sandoval [24] analyzes both the MST as well as the network with several thresholds.

Several studies have been published on a network analysis of a specific stock market, in particular on the New York Stock Exchange [25,26,11,27] but also on other countries' exchanges [18,28,29,21,15,30,10,8]. Due to its small capitalization and to scarcity of historical data [31], the Italian stock market has very few analyses, in particular by Brida and Risso [32,33], who used Yahoo!'s data for a set of a relatively small number of companies and short time. Few other authors have built alternative networks for the Italian stock market, based on boards of directors [34] or ownership [35]. The interesting innovation introduced by Brida and Risso and other authors [36,37] is the symbolization of data, which is able to filter out the micro-movements typical of stock prices [38]. Its main useful application is, however, giving the possibility to simultaneously consider prices and other quantitative data which use different measures, such as volumes or even non-daily data such as dividends or yearly accounting data. In fact, introducing volumes in price movement analysis is a common practice among professionals. Karpoff finds a positive asymmetric relationship between volume and price change [39–41], Stošic et al. [42] find relationships between price and volume returns, suggesting their complex dynamics come from both intrinsic properties of the series and mutual interactions. Lo and Wang [43] confirm the importance of trading volume in modeling asset markets. The effect of these additional data can be introduced by simply multiplying price or return by volume [44] or calculating the RV coefficient [45], but these strategies do not take full benefit of symbolization. Further they can be introduced appending the symbolized volumes to the time-series of the symbolized returns when calculating the correlation matrix or even combining volumes and returns using symbolization rules which reflect their financial meaning. Few authors analyze the different methods and techniques to build stock market networks, for example Ref. [4] compare different time frames, [10] compare different clustering techniques, [12] analyze the introduction of cliques in minimum spanning trees, but we found no comparison for MSTs obtained with symbolization.

This work presents four network structures of the Italian stock market in the 2001–2011 period built with different symbolization methods. The first one is Mantegna's standard method [3] using the correlations of log-returns to build a distance matrix among all the 100 involved companies and consequently a network, from which we extract a minimum spanning tree using Kruskal's algorithm. The second one replaces the log-returns by three symbols and uses them to build the correlations, filtering out the noise due to small fluctuations typical of stock markets. Our third method is the same as the second one but extending the series of symbolized log-returns with a series of symbolized traded money, thus obtaining longer series on which correlations can be calculated. The last method still uses symbolized log-returns and symbolized traded money but combining them according to the idea that a large price variation must be confirmed by a large amount of traded money to be taken into consideration.

Unlike [33] we use many more companies and we take data with a better quality, as they have been manually checked in particular regarding volumes and corporate actions such as mergers and takeovers. Section 1 of the article presents the starting raw data source and the manipulations to obtain a return free of spurious effects from dividends or other corporate actions, which can artificially influence the price.[1] Section 2 proceeds with symbolization and the building of the correlation matrix using a metric distance. The next illustrates the building of the four MSTs and their reliability measures and presents the definition of the measures which will be used to summarize and compare the four methods. Section 4 shows the results for the four MSTs and their measures. Finally, the conclusions sum up the most interesting results and open new scenarios for future developments.

[1] When a company pays a dividend, its share price artificially drops by approximately the dividend's amount. When a company increases its capital, the value of the current shares increases thanks to the new fresh money flowing into the company and at the same time it is diluted due to the arrival of other shares with dividend and voting rights.