Physica A 462 (2016) 410–420

Contents lists available at ScienceDirect

Physica A

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

Clustering stocks using partial correlation coefficients

Sean S. Jung, Woojin Chang [*]

Department of Industrial Engineering, Seoul National University, Republic of Korea

HIGHLIGHTS

- Correlation analyses are conducted on Korean stock market.
- Agglomerative hierarchical clustering is performed based on correlation matrices.
- Each cluster consists of firms from multiple business sectors.

ARTICLE INFO

Article history:
Received 15 October 2015
Received in revised form 14 May 2016
Available online 23 June 2016

Keywords:
Financial market
Stylized facts
Pearson correlation
Partial correlation
Agglomerative clustering

ABSTRACT

A partial correlation analysis is performed on the Korean stock market (KOSPI). The difference between Pearson correlation and the partial correlation is analyzed and it is found that when conditioned on the market return, Pearson correlation coefficients are generally greater than those of the partial correlation, which implies that the market return tends to drive up the correlation between stock returns. A clustering analysis is then performed to study the market structure given by the partial correlation analysis and the members of the clusters are compared with the Global Industry Classification Standard (GICS). The initial hypothesis is that the firms in the same GICS sector are clustered together since they are in a similar business and environment. However, the result is inconsistent with the hypothesis and most clusters are a mix of multiple sectors suggesting that the traditional approach of using sectors to determine the proximity between stocks may not be sufficient enough to diversify a portfolio.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

For decades, the financial market has received an enormous amount of attention from academia. Yet its complex nature still remains elusive and recent financial crises support the need to better understand it. Statistical Physics is one of the popular tools to analyze financial data and many important discoveries have been made by using it. For example, though a financial asset's returns show no serial correlation, the absolute values of returns have positive correlation over long lags, a phenomenon well known as the long memory property [1]. Also, a financial return series has heavier tails than Gaussian distribution and the absolute value of the return series tends to follow a power law [2,3]. It also shows nonlinearity in behavior as observed in phenomena such as volatility clustering and regime switching behavior [4–7]. Another interesting finding is that when multiple assets are considered together, the return series are often correlated and the correlation tends to become larger during a financial crisis [8,9]. Correlation plays a critical role in analysis of financial time series. For example, it has been known for a while that the correlation in equity returns is time varying [10]. Also, co-movement of international markets was widely studied and methods such as conditional correlation and dynamic correlation were some of the most

---

[*] Corresponding author.
E-mail address: changw@snu.ac.kr (W. Chang).

http://dx.doi.org/10.1016/j.physa.2016.06.094
0378-4371/© 2016 Elsevier B.V. All rights reserved.

popular tools to approach this subject [11–13]. Network theory can be applied to study a correlation matrix as well, and previously it was demonstrated that a correlation matrix can be associated with a hierarchical tree and when tested in a small set of stocks, the stocks in the same sectors were clustered together [14]. When it comes to the construction of portfolio of assets, correlation is one of the key variables an investor needs to consider [15,16].

This study employs the correlation to study how firms are related to others over different periods of time including the recent financial crises. Pearson correlation is widely accepted as a standard measure of co-movement between two financial return series but it has few weaknesses. For example, as it was mentioned before, the market dynamics are nonlinear in nature and so are the correlations between assets, but Pearson correlation measures a linear co-movement between two random variables. Also, when it comes to analysis of firms in a similar environment, one should consider the common factors affecting the firms such as the growth rate of market or foreign exchange rates. Those common factors may have an influence on the stocks, resulting in a bias in analysis. Such bias may skew the results to favor one side to the other, giving it a 'false correlation'. Though it is impossible to completely remove all external factors from a financial time series as some of the factors are unobservable or difficult to assign numerical values such as the corporate governance, it is likely that removing some of the obvious and prevalent factors may provide different insights which have not discovered.

Recently, Kenett et al. applied the concept of partial (residual) correlation analysis to study the financial markets [17]. They successfully demonstrated the index cohesive effect and identified a dominating sector within a market using the partial correlation analysis [18,19]. The network approach was also used to construct node–node correlation matrices and it was shown that a node with insignificant influence does not disrupt the network even when the node is removed [20]. Another study proposed a measure called Sector Dominance Ratio using Pearson and partial correlations to study the market structure and it has empirically shown that the financial sector exhibits strong dominance for US and UK markets [21]. The studies mentioned above utilize the partial correlation analysis and the strength of this method comes from its ability to remove a common factor in correlation between two variables. A partial correlation measures how a random variable '*i*' correlates with another random variable '*j*' when a common factor '*k*' is removed from both of them. In this sense, it can be said that it is correlation between residuals of *i* and *j*.

One may notice that if a common factor is chosen to be the market return, the returns used to compute correlation resemble the return from Capital Asset Pricing Model [22–24]. CAPM was developed to describe how a risky asset should be valued and if the model was correct, the correlation between the residuals, or the partial correlation coefficient of two risky assets, should be statistically zero. Section 3 discusses this point and shows that the correlation coefficients are in fact non-zero. This should not come as a surprise as the CAPM was known for its relatively weak explanatory power [25].

This study is another application of the partial correlation analysis. A market index such as Korea Composite Stock Price Index (KOSPI) reflects the overall size of Korean economy. A growth of the index means that the size of economy is growing as well which will positively affect every firm in the market. By removing the effect of the market index, it may be possible to shed a new light regarding the individual performance of the stocks and their correlation with others. Specifically, it addresses following two questions:

i. How much difference exists between the partial correlation and Pearson correlation? Does the difference persist through different period of time?
ii. In terms of the partial correlation, how are firms related to each other?

Both the partial correlation and Pearson correlation are computed every 30 months to minimize the effect of change in correlation coefficients which occurs due to the nonlinearity of the market. The length of 30 months also ensures that the computed correlation coefficients are statistically significant. It is then visualized to address the first question and a simple agglomerative clustering approach is used to explore the second question.

The remaining sections of this paper are organized as follows: Section 2 describes the data set used for analysis and method of how to compute the correlations used in this paper. Section 3 reports the results of the partial correlation analysis and Pearson correlation analysis. In Section 4, clustering analyses are performed. Section 5 provides the conclusion.

2. Data and methods

The monthly adjusted closing price series of KOSPI and the stocks listed in the index from December of 2004 to December of 2014 are used for this study. The data are provided by DataGuide from FnGuide (http://www.fnguide.com/), a professional financial analytics service for the Korean market. All computation and analytics are done using MATLAB. Total of 732 firms existed throughout this period and trade volume data are used to filter out the ones that did not trade during a period of analysis. Note that not all the firms existed for the entire period of time resulting in different number of firms and different size of correlation matrices for each period.

For each firm, a log return $r_i(t)$ was computed by

$$r_i\left(t\right) = \ln\left(P_i\left(t\right)\right) - \ln\left(P_i\left(t-1\right)\right) \tag{1}$$

where $P_i(t)$ denotes the adjusted closing price of firm *i* at time *t*. The log return series are used for analyses which covered ten years from January of 2005 to December of 2014.

Download English Version:

https://daneshyari.com/en/article/7377130

Download Persian Version:

https://daneshyari.com/article/7377130

Daneshyari.com