Detecting Correlation in Stock Market

Jörg D. Wichard, Christian Merkwirth, Maciej Ogorzałek [a,b,a]

[a] *AGH University of Science and Technology*
Department of Electrical Engineering
al. Mickiewicza 30
30-059 Kraków, Poland

[b] *Max-Planck-Institut für Informatik*
Stuhlsatzenhausweg 85
66123 Saarbrücken, Germany

Abstract

We present a new method for detecting dependencies in the stock market. In order to find hidden correlations in the daily returns, we build cross prediction models and use the normalized modeling error as a generalized correlation measure that extends the concept of the classical correlation matrix.

Key words: Econophysics, Multivariate analysis, Time series analysis
PACS: 89.65.Gh, 02.50.Sk, 05.45.Tp

1 Introduction

The analysis of the the cross-correlation matrix of the returns plays an important role in portfolio theory and financial analysis. We build the time series of daily returns

$$R_i(t) = \frac{Y_i(t+1) - Y_i(t)}{Y_i(t)},$$

wherein $Y_i(t)$ denotes the closing-price of the i-th stock at day t. The cross-correlation matrix of the returns is defined as

$$\rho_{ij} = \frac{\langle R_i R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{\langle R_i^2 - \langle R_i \rangle^2 \rangle \langle R_j^2 - \langle R_j \rangle^2 \rangle}},$$

where the brackets indicate the time average over all trading days in the investigated period. The analysis of ρ_{ij} leads to some interesting insights in the market dynamics. Mantegna (see Mantegna (1999)) discovered a hierarchical

Preprint submitted to Elsevier Science *25 April 2004*

organization inside a portfolio of stocks by introducing a metric related to the correlation coefficients. By definition the correlation matrix is symmetric with respect to i and j and thus cannot be used to distinguish a symmetrical interaction between different stocks from an asymmetric one. Following our investigations we see strong indications that this asymmetric interaction exists in a way that the dynamics of single stocks are leading the dynamics of others significantly. We indicate this with a cross modeling scheme which is described in the following section.

2 Mixed State Analysis

The scheme we introduce for market analysis is related to the "mixed state analysis" of multivariate time series which was developed to detect weak coupling between dynamical systems in the framework of chaotic synchronization (see Wiesenfeldt et al. (2001)). This approach is based on the reconstruction of mixed states consisting of delayed samples taken from simultaneously measured time series of both systems under investigation.
We adopted this idea and changed it for our purpose in a way that a linear model $f(\vec{\mathbf{R}}_{i,j}(t))$ is constructed that maps the time-lagged returns of the j-th stock together with the time-lagged returns of the i-th stock

$$\vec{\mathbf{R}}_{i,j}(t) = (R_j(t), R_j(t-1), \ldots, R_j(t-\tau),\ R_i(t-1), \ldots, R_i(t-\tau)) \qquad (1)$$

onto the actual returns of the i-th stock $R_i(t)$. The model $f(\cdot)$ is a linear function that is fitted using the standard least squares approach (see for example Hastie et al. (2001)) for multiple linear regression models, i.e. it should minimize the residual sum of squares $\sum_t (R_i(t) - f(\vec{\mathbf{R}}_{i,j}(t)))^2$. We would like to remark that this model $f(\cdot)$ is for sure not able to make *predictions* of the returns for the next day, however it is able to find the relationship between the actual returns $R_i(t)$ and $R_j(t)$ with respect to the time lagged returns, that may contain some information about linear trends on short time scales.
If we consider a portfolio of N different stocks, we can define the $N \times N$-matrix of the normalized modeling error as

$$cp(i,j) = \frac{\langle (R_i - f(\vec{\mathbf{R}}_{i,j}))^2 \rangle}{\langle R_i^2 - \langle R_i \rangle^2 \rangle}, \qquad (2)$$

where the brackets denote the time average. The modeling error is normalized with the variance of the time series $R_i(t)$ for a simple reason: A value of $cp(i,j) \geq 1.0$ indicates that the mean value $\langle R_i \rangle$ is a more appropriate model than $f(\cdot)$, which means that there is no linear dependence in the the time series under investigation. Smaller values of $cp(i,j)$ give an indication that there is at least a weak linear interrelation between the dynamics of the returns. In general, the matrix $cp(i,j)$ is not symmetric, i.e. $cp(i,j) \neq cp(j,i)$. We define the matrix of differences $\delta(i,j)$ as

$$\delta(i,j) = cp(i,j) - cp(j,i). \tag{3}$$

The values of $\delta(i,j)$ reflect asymmetric dependencies in the market dynamics. If the returns of i and j are uncorrelated or they interact on the same level, then we expect $\delta(i,j) \approx 0$.
For $\delta(i,j) > 0$ we have $cp(i,j) > cp(j,i)$ which means that the returns of the i-th stock contain more useful information to model the returns of the j-th stock than the other way around. In the terms of synchronization this indicates an asymmetrical coupling strength between the two stocks.

3 Numerical Simulations

We investigate 600 trading days of the Dow-Jones Industrial Average (DJIA) between 2-Oct-2000 and 3-Mar-2003. For all 30 stocks in the DJIA, we build the time series of daily returns and calculate the cross-correlation matrix $\rho(i,j)$ (see equation 1). For the mixed state analysis we use a time lag of $\tau = 3$ and we calculate the matrix of the modeling error [1] as defined in equation 2 and further the matrix of differences $\delta(i,j)$ from equation 3. The results are shown in Figure 2. The cross-correlation matrix shows some interesting structures, for example are there obvious clusters, there were described by Mantegna (1999). A part of this structures can be found in the matrix of the modeling error $cp(i,j)$. The stocks that behave anti correlated with respect to the index (the blue stripes in the correlation matrix) occur in $cp(i,j)$ with an modeling error near one. In the matrix of the error differences $\delta(i,j)$ we find the amount of asymmetry regarding our mixed state analysis that offers a field of further investigations. The next step will be a detailed analysis of the time dependence of these asymmetries an the nonlinear dependencies in the stock market.

References

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer Series in Statistics. Springer-Verlag.

Mantegna, R., 1999. Hierarchical structure in financial markets. Eur. Phys. J. B. 11, 193–197.

Wiesenfeldt, M., Parlitz, U., Lauterborn, W., 2001. Mixed state analysis of multivariate time series. Int. J. Bifurcation and Chaos 11 (8), 2217–2226.

[1] In order to achieve a better graphical resolution in the plots, we set the zero diagonal elements to one.

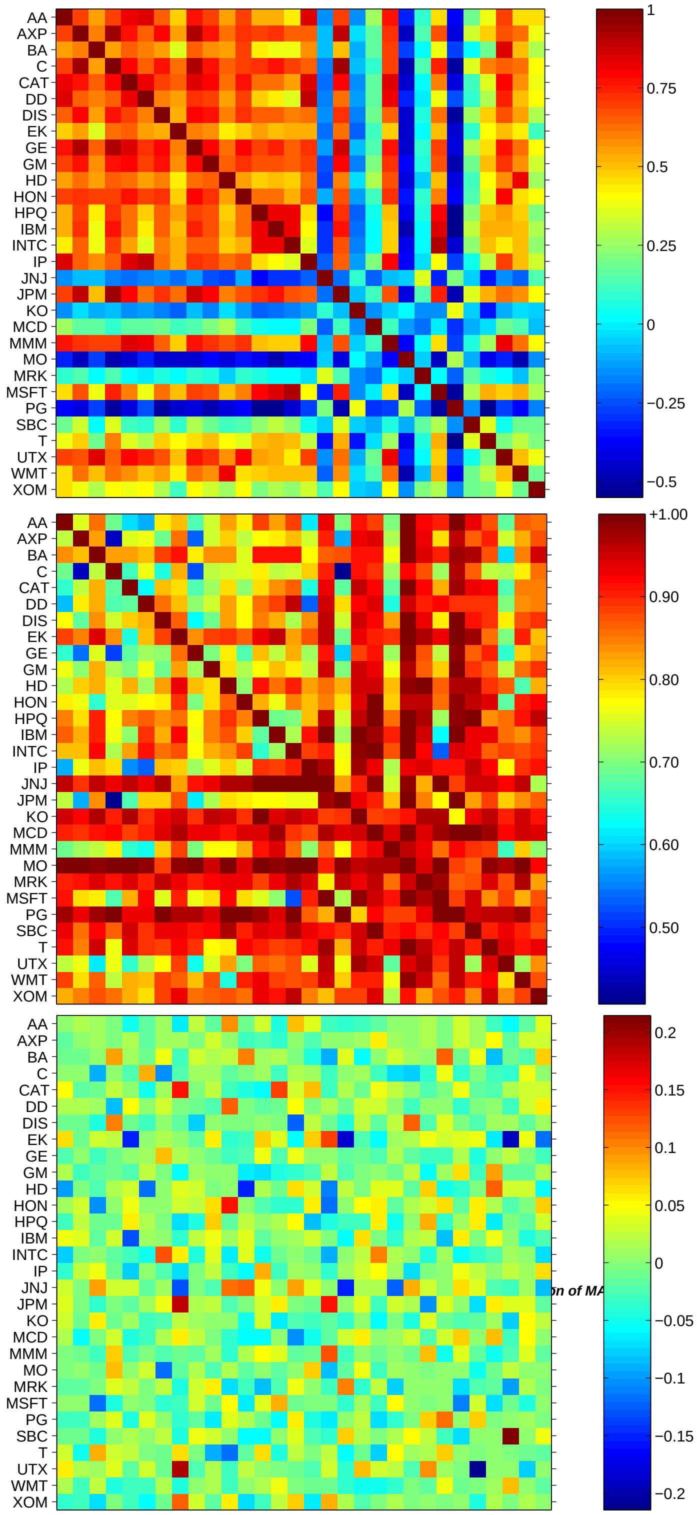

Fig. 1. The cross-correlation matrix (top), the matrix of the normalized modeling error $cp(i,j)$ (middle) and the matrix $\delta(i,j)$ of the error differences as defined in equation 3 (bottom) for 600 days of the DJIA (Ticker symbols on the left).