DETECTING CORRELATION IN STOCK MARKET WITH CROSS PREDICTION

J.D. Wichard and M.J. Ogorzałek

AGH - University of Science and Technology
al. Mickiewicza 30
30-059 Kraków, Poland

Correlation in Stock Markets

We present a method for detecting correlations in the stock market based on a nonlinear modeling approach. In order to find dependencies in the daily returns, we build nonlinear cross prediction models and use the prediction error as a generalized correlation measure, that extends the classical correlation matrix. The analysis of the the cross-correlation matrix of the returns plays an important role in portfolio theory. Therefor 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 cross-correlation matrix is symmetric by definition, so it might not reflect asymmetric dependencies in the market.

Cross Prediction

We like to introduce a cross-prediction scheme, to detect correlations and dependencies in the stock market. It 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 [1].
We fit a nonlinear model to predict the daily returns $R_i(t)$ as a function of the time-lagged returns of the j-th stock $\vec{\mathcal{R}}_j(t) = (R_j(t), R_j(t-1), \ldots, R_j(t-\tau))$ so we have

$$R_i(t) = f(\vec{\mathcal{R}}_j(t)). \tag{1}$$

The model $f(\cdot)$ is for sure not able to make precise predictions of the desired returns, but it is able to detect dependencies in the stock market, if we analyze the resulting modeling error. We define the normalized cross-prediction error as

$$cp(i,j) = \frac{\langle (R_i - f(\vec{\mathcal{R}}_j))^2 \rangle}{\langle R_i^2 - \langle R_i \rangle^2 \rangle}, \tag{2}$$

where the brackets denote the time average. If $cp(i,j)$ is close to 1.0 this means, that the model cannot find any relation between the two stocks i and j. A value smaller than 1.0 indicates a dependence and the difference $cp(i,j) - cp(j,i)$ contains information about the lack of symmetry between the time evolution of the considered stocks.

The work was done within the Research Training Network *COSYC of SENS* No. HPRN-CT-2000-00158 within in 5th EU Framework Program of the European Community.

Results

We investigate 600 trading days of the Dow-Jones Industrial Average (DJIA) between 2-Oct-2000 and 3-Mar-2003. We divided the data in two sets of 300 days each. The first set was used to train the nonlinear models as defined in Equation 1. We trained feed forward neural networks with two neurons in a single hidden layer for each pair of stocks and calculate the normalized cross-prediction error on this training set, called the "in-sample-error". Then we took the data that we didn't use for model-training and calculated again the normalized cross-prediction error, but this time "out-of-sample" (OOS). The results are shown in Figure 1. The OOS-error is slightly larger than in the training set, which is natural, but the structure of the matrix is preserved, which shows that we found real dependencies in the market, that could be used for further investigation.

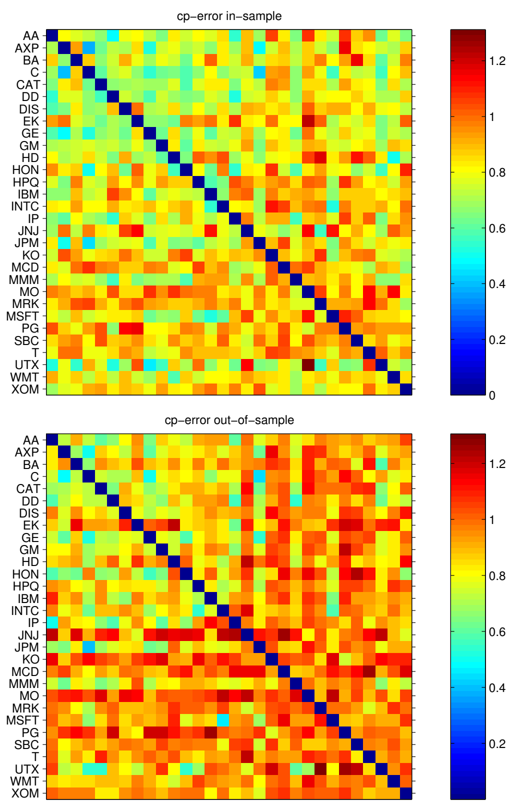

Figure 1: The value of $cp(i,j)$ for the 30 stocks in the DJIA (Ticker Symbols) for the training set (in-sample) and the test set (out-of-sample).

References

[1] M. Wiesenfeldt, U. Parlitz and W. Lauterborn, Mixed State Analysis of Multivariate Time Series. Int. J. Bifurcation and Chaos **11** (8), 2217-2226 (2001).