

固有値分析とポートフォリオ構築

Eigenvalue Analysis and Portfolio Optimization in Financial Markets

上田唯以¹ 橋本康弘¹ 陳 Yu¹ 大橋弘忠¹Yui Ueda¹, Yasuhiro Hashimoto¹, Yu Chen¹, Hirotada Ohashi¹¹ 東京大学大学院工学系研究科システム創成学専攻¹ Department of Systems Innovation Graduate School of Engineering, The University of Tokyo

Abstract: Recently, a lot of studies on the eigenvalue analysis are performed in order to investigate the statistical characteristics of multi-assets correlation in financial markets. In this study, we analyzed the eigenvalue of the cross-correlation matrix of the stock prices listed in the Tokyo Stock Exchange. A filtered cross-correlation matrix is built by removing the noise mode. Comparing the network graphs visualized from normal cross-correlation matrix with the one from the filtered cross-correlation matrix, the latter reflects the characteristic of the market in a more insightful manner. In addition, we build a portfolio from the filtered cross-correlation matrix, and carry out a backtesting by doing a simulated investment. As a result, the portfolio made from the filtered cross-correlation matrix shows better performance than the normal one. Our result suggests that eigenvalue analysis is useful for both understanding the market structure and improving the portfolio for investment.

1 はじめに

様々な金融データを対象として計算機等を利用した分析が積極的に行われるようになった。近年、複数銘柄を対象とした研究の一つとして、株式市場の相関行列の固有値分析が行われている。([1], [2], [3]) 相関行列の固有値分析を行うことで一定の特徴が観察され、それらの特徴がランダム行列理論で説明できるなど、相関行列のランダム性に関して解明が進められ、モデル化による近似などもされるようになってきた。一方で、相関行列の固有値の中にはランダム行列理論だけでは説明しきれない部分などもあり、株式市場における相関行列の理解を深めるためにも、更なる研究が期待されている。本研究では、Pan ら [4] の研究である固有値による相関行列のフィルタリングを利用し、日本の株式市場（東京証券取引所）のネットワーク分析とポートフォリオの構築を行った。ネットワーク分析では、フィルター化された相関行列ネットワークを用いることで、特定の業種間の相関が抽出できうることを述べ、ポートフォリオ構築では、フィルター化した相関行列を用いたポートフォリオ構築と従来の相関行列のポートフォリオとに関して運用のバックテストを行い、フィルター化した相関行列のポートフォリオが良好な運用結果を残したということを述べる。

2 研究手法

本節では本研究で用いた手法などの説明を行う。なお、詳しい数理的な部分は割愛し、本質的な部分のみの解説にとどめることとする。

2.1 相関行列と固有値

$P_i(t)$ を時刻 $t (= 1, 2, \dots, T)$ における銘柄 $i (= 1, 2, \dots, N)$ の証券価格としたとき、銘柄 i のログリターン $R_i(t)$ を以下のように定義する。

$$R_i(t) \equiv \ln P_i(t+1) - \ln P_i(t) \quad (1)$$

ログリターンをその銘柄のリスク（標準偏差）によって標準化したリターンを $r_i(t)$ とすると以下のようになる。

$$r_i(t) \equiv \frac{R_i - \langle R_i \rangle}{\sigma_i} \quad (2)$$

ただし、 $\sigma_i \equiv \sqrt{\langle R_i^2 \rangle - \langle R_i \rangle^2}$ であり、 $\langle \dots \rangle$ はその期間における平均を表す。この標準化したリターンを用いて、相関行列 C における銘柄 i と j の成分 C_{ij} を以下のように定義する。

$$C_{ij} \equiv \langle r_i r_j \rangle \quad (3)$$

相関行列 C の固有値を降順にしたものを $\lambda_i (i = 1, 2, \dots, N)$, その固有値に対応する固有ベクトルを $\mathbf{u}_i (i = 1, 2, \dots, N)$ とすると、

$$C = \sum_{i=1}^N \lambda_i \mathbf{u}_i \mathbf{u}_i^T \quad (4)$$

とすることができる。ただし、 \mathbf{u}_i^T は \mathbf{u}_i の転置行列である。また、いくつかの銘柄が固有ベクトルの要素として大きく関わっているかを測るものとして Inverse participation ratio (IPR) を以下のように定義する。

$$I_k \equiv \sum_{i=1}^N u_{ki}^4 \quad (5)$$

u_{ki} は k 番目の固有値に対応する固有ベクトルにおける、銘柄 i の値である。IPR は u_{ki} の一つが 1、他が 0 のときに最大値の 1 となり、 u_{ki} の全ての要素が $1/N$ であるときに最小値の $1/\sqrt{N}$ となる。IPR は大きいほど一部の要素の寄与が大きく、小さいほど均等に寄与していることを示す。なお、ランダム行列によって得られる固有値の分布は以下の式で求められる。

$$P_{rm}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda} \quad (6)$$

2.2 固有値を利用した相関行列のフィルタリング

相関行列の要素が、マーケットによる要素、グループによる要素、ランダムの要素という三要素から成り立っていると仮定すると、(4) 式を利用して相関行列をフィルタリングすることができ、

$$C = C^{\text{market}} + C^{\text{group}} + C^{\text{random}} \\ = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \sum_{i=2}^{N_g} \lambda_i \mathbf{u}_i \mathbf{u}_i^T + \sum_{i=N_g+1}^N \lambda_i \mathbf{u}_i \mathbf{u}_i^T \quad (7)$$

N_g は固有値や固有ベクトルの特徴などを基に決定する。決まった値はなく、証券市場や対象銘柄の取り方などによって異なる。

2.3 ネットワーク分析

相関行列 C や 2.2 で得られたフィルタ化相関行列などを基に、ネットワークを作成することができる。本研究では、 C 、 C^{market} 、 C^{group} を対象とし、単純に閾値 c^{th} を定め、それぞれの要素が c^{th} より大きければ 1 (接続)、小さければ 0 (非接続) とすることでネットワークを作成した。得られたネットワークのデータを s.o.c.i.a.r.i.u.m [5] というソフトを用いて可視化した。

2.4 フィルタ化分散共分散行列の算出

ポートフォリオの構築に関しては Markovitz [6] をベースにして行う。ポートフォリオ構築には分散共分散行列 Σ を得る必要があり、それは相関行列を基に以下のようにして求めることができる。

$$\Sigma = \sigma C \sigma \quad (8)$$

ただし、 σ は、対角成分が銘柄 i の標準偏差 σ_i になっている対角行列である。今回は、(8) 式における C を、フィルタ化した相関行列 ($C^{\text{market}} + C^{\text{group}}$ 、以下 C^{mg}) に置き換えることでフィルタ化した分散共分散行列 Σ^{mg} を算出する。すなわち、

$$\Sigma^{\text{mg}} = \sigma C^{\text{mg}} \sigma \quad (9)$$

という式によって算出できることとなる。

2.5 ポートフォリオ構築

リターンの時系列データと分散共分散行列さえあれば、最適なポートフォリオを求めることが可能である。ただし、解析解を求めるためには、分散共分散行列が正定値行列でなければならない。通常の分散共分散行列 Σ は正定値行列になるのだが、フィルタ化分散共分散行列 Σ^{mg} は、正定値行列にならないこともある。(ただし、論理的に証明したわけではなく、実際にデータを扱ってみたところ正定値行列にならなかったにすぎない。) 従って通常の解析解は得られないため、本研究では遺伝的アルゴリズムを用いて確率解を求め、その確率解を基にポートフォリオを構築した。

2.6 遺伝的アルゴリズム

本研究における遺伝的アルゴリズムの設定は以下の通りになる。

- 世代あたりの個体は 100
- 個体の持つ要素は、ポートフォリオの各銘柄のウェイトで合計は 1。解として求めたいのはこのウェイト
- 繰り返し回数は 10000 もしくは 100000
- 交叉は一様交叉で、確率 0.5 で個体同士の要素を入れ替える。交叉に選ばれる個体はランダム
- 突然変異はなし
- 選択はエリート選択で個体は 2 個だけ残す

- 評価関数は、(ポートフォリオのリターン) ÷ (ポートフォリオの標準偏差)
- 初期世代は、100 個体全てが Σ を用いて求めたポートフォリオの最適解で、解を求めるときの期待リターンは各銘柄のリターンの平均

この設定の下、以下の手順で世代を進めていく。

1. 現世代の 2 個体を選択して交叉を行い、新世代を生成する。
2. 現世代と新世代に関して評価関数を算出する
3. 評価関数の最も高い 2 個体を選択する
4. 残った 2 個体のうち、どちらかをランダムに選択する
5. 選択された個体の要素のどれか一つに、一様乱数 (0 ~ 0.001) を加えた後に規格化してウェイトの合計を 1 に戻す。これを新たな個体とする
6. 4. ~ 5. の作業を世代の個体数が 100 になるまで繰り返す

繰り返しが最後まで到達した時点で最も評価関数が高い個体の要素をポートフォリオの解とする。

2.7 運用シミュレーション

フィルター化相関行列を用いたポートフォリオ構築が有用であるかを検証するため、単純なバックテストを行う。データの期間などは後述する。バックテストの設定としては、

- 前 300 営業日分のデータを用いてポートフォリオを構築する
- 構築したポートフォリオで 50 営業日運用する
- 50 営業日運用した後、再び前 300 営業日分のデータを用いてポートフォリオを構築する
- 上記を 16 回行う。すなわち、800 営業日に渡って運用を行うことになる
- ショートポジションは不可、取引に伴う手数料、最小取引量の制限などはなし
- 対戦相手は通常の相関行列を用いたポートフォリオの解析解。ただし、解析解の期待リターンは遺伝的アルゴリズムによって得られた期待リターンと等しくする。

運用結果は、インデックス (今回は日経平均株価) に対する超過収益で評価する。リターンがプラスでも、インデックスよりもリターンが悪ければマイナス評価になるということである。

2.8 データに関して

固有値分析は東京証券取引所における 2004 年 1 月 5 日から 2008 年 7 月 14 日までの日足データを用いた ($T = 1114$)。対象とする銘柄は、2008 年 7 月において日経平均株価の算出に使用されている銘柄のうち、対象期間においてデータの欠損の無い 211 銘柄である ($N = 211$)。運用バックテストは、実際の運用期間が 2005 年 3 月 25 日から 2008 年 6 月 24 日になる。すなわち、2004 年 1 月 4 日から 2005 年 3 月 24 日までのデータが最初のポートフォリオ構築に用いられることになる。データは「パンローリング相場データ集国内相場版」を利用した。

3 結果・考察

本節では本研究で得られた結果を示し、それに関する考察を述べる。

3.1 固有値分析

データを基に固有値と IPR を求め、グラフにしたものが図 1 である。

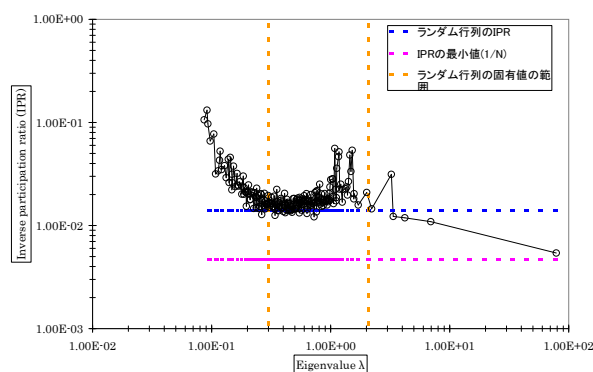


図 1: 固有値と IPR のグラフ

図中の青線はランダム行列の IPR の平均値、赤線は IPR の最小値、橙色の線はランダム行列の固有値の範囲を示す。図 1 より、大部分の固有値がランダム行列の範囲内にあるが、最大固有値は約 79 でランダム行列の範囲から大きく乖離 ((6) 式の最大値のおよそ 40 倍) していることが分かる。また、最大の固有値からランダム行列の範囲に入るまでの固有値の数が 5 ~ 6 ほど観察されるため、日本市場における N_g の値は 5 ~ 6 であると推測される。

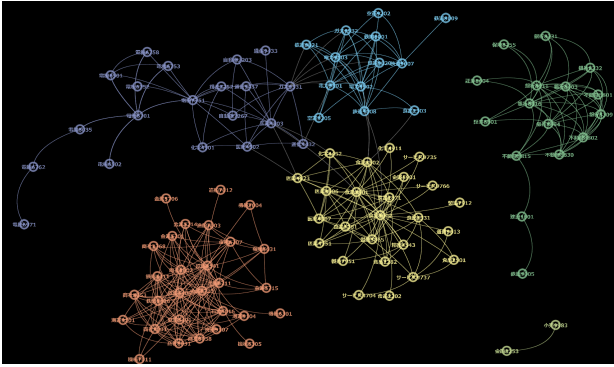


図 2: Cgroup のネットワーク可視化

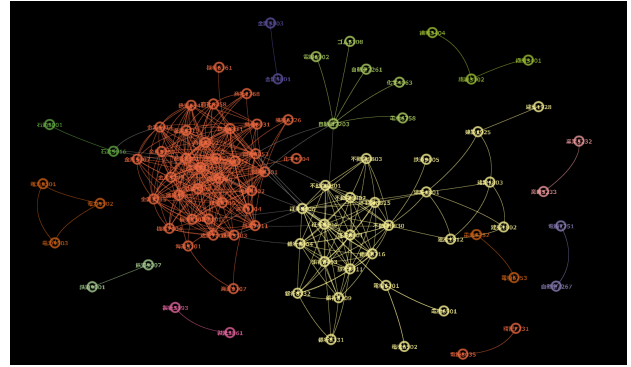


図 4: C のネットワーク可視化

3.2 ネットワーク分析

N_g の値が 5~6 であることが分かったため、(7) 式を用いて相関行列のフィルタリングを行うことができる。フィルタリングの後、 C^{group} からネットワークを作成し、可視化したものが図 2 である。閾値である c^{th} は 0.12 に設定し、色づけはモジュラリティー最適化によるクラスター分けを基に行っている。

一見ただけでいくつかのグループに分かれていることが把握できる。さらに詳しくネットワーク中の銘柄を見ていくと、左上の紫色のグループは自動車・電機・通信など、中央上の水色のグループは電力・ガス・鉄道・空輸など、右の緑色のグループは保険・銀行・証券など、左下の橙色のグループは鉄鋼・造船・商社・海運など、中央下の黄色のグループは医薬・化学・食品などといったように、産業の分野ごとのグループになっていることが分かった。このことから、 C^{group} は相関行列の要素のうち、グループごとの相関特性が含まれているものではないかと考えられる。

続いて、 C^{market} からネットワークを作成し、可視化したものが図 3 である。 c^{th} は 0.55 に設定している。 C^{group} の場合と異なり、一見できるほどのグループ

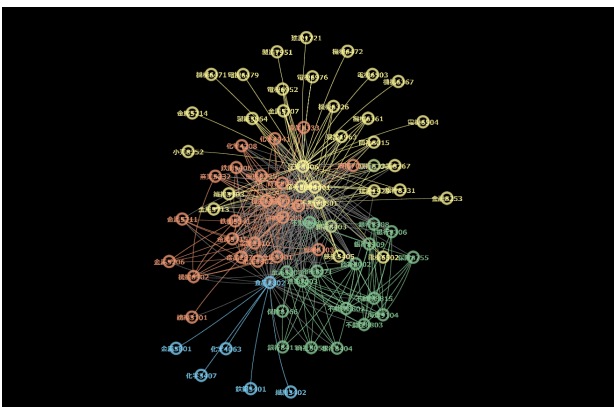


図 3: Cmarket のネットワーク可視化

分けはできておらず、多くの銘柄が固まっているように見える。さらに、色別のグループごとに銘柄を見ても C^{group} のときのように特徴的な分類がされていない。このことから、 C^{market} は相関行列の要素のうち、マーケット全体が上下することによって生み出される相関特性を示しているのではないかと考えられる。

最後に、相関行列全体である C のネットワークを作成し、可視化したものが図 4 である。 c^{th} は 0.60 に設定している。

C のネットワークは、大きな塊のグループが少数で、2~3 程度の銘柄が繋がった小グループが多数できている。これは、C そのものが C^{market} 、 C^{group} 、 C^{random} で成り立っているとすれば、大きな塊のグループは C^{market} が由来となって出来ており、小さなグループは C^{group} が由来となって出来ていると考えることができる。複数の要素が合成されているため、 C^{market} や C^{group} のみのネットワークに比べ、はっきりとした特徴が捉えにくくなっている上に、 C^{random} の影響によって、さらに特徴が分かりにくくなっている可能性があると考えられる。

3.3 運用シミュレーション

遺伝的アルゴリズムのループ回数を 1 万回に設定したときの運用シミュレーションの結果が図 5 である。図 5 はインデックスに対する超過収益であるため、運用結果がマイナスであっても、絶対リターンがマイナスというわけではない点に注意されたい。(逆もそうである。) 運用結果をまとめたものが表 1 である。

	log リターン	リターン	標準偏差
フィルター相関	0.136	14.6%	10.8%
従来の相関	-0.067	-6.5%	8.4%

表 1: 運用結果 (ループ 1 万)



図 5: 運用結果比較 (ループ 1 万)

フィルター化相関を用いた運用は、従来の相関を用いた運用と比べて、リターンが約 20% 上昇しているにも関わらず、標準偏差が約 2% しか上昇しておらず、運用成績が向上していると考えられる。また 50 営業日を 1 期とし、各期ごとの運用成績をみたところ、16 期中 11 期で運用成績が勝っており、安定してよい運用結果を出していることが分かる。しかし、(年率超過リターン)/(標準偏差)で求められる Information Ratio に関しては 0.4 程度に過ぎないため、絶対的に運用の成績がよいというわけではない。

次に、遺伝的アルゴリズムのループ回数を 10 万回に設定して運用シミュレーションを行った。その結果が図 6 と表 2 である。

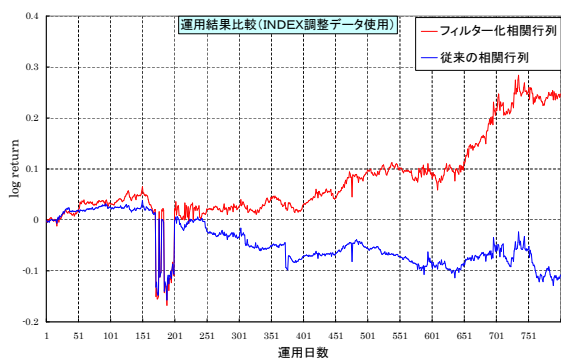


図 6: 運用結果比較 (ループ 10 万)

	log リターン	リターン	標準偏差
フィルター相関	0.247	28.0%	8.96%
従来の相関	-0.108	-10.2%	4.41%

表 2: 運用結果 (ループ 10 万)

1 万回の結果と比較すると、全体的に分散が減少し、最終的な運用成績は向上しており、16 期中で 15 期に渡って運用成績が勝っている。ただし、途中段階における運用成績に関しては 1 万回の結果に比べて悪化しており、運用成績に関する試行回数の影響は、選択する銘柄や運用期間などによって様々に変化すると考えられる。なお、従来の相関行列の運用結果が変化したのは、遺伝的アルゴリズムのループ回数が増えることにより、フィルター化相関行列のポートフォリオの期待リターンが変化したことに伴い、従来の相関行列の期待リターンも変化したからである。

今回のシミュレーション条件に関して言えば、試行回数が 1 万回にしる 10 万回にしる、フィルター化相関行列を用いることで運用成績は向上したといえる。その理由に関しては様々なものが考えられるが、理由の一つとして、従来の相関行列にランダム性 (ノイズ) が多く含まれることが挙げられる。図 7 は、相関行列の要素の値の分布をフィルターごとにみたものであるが、これを見ると、 C^{group} の値とほぼ同じ値を C^{random} が持っていることがわかる。すなわち、 C^{random} の影響によって、 C^{group} の情報が希薄化されてしまい、結果としてポートフォリオの最適化の際のノイズの影響が大きくなってしまっていると考えられる。

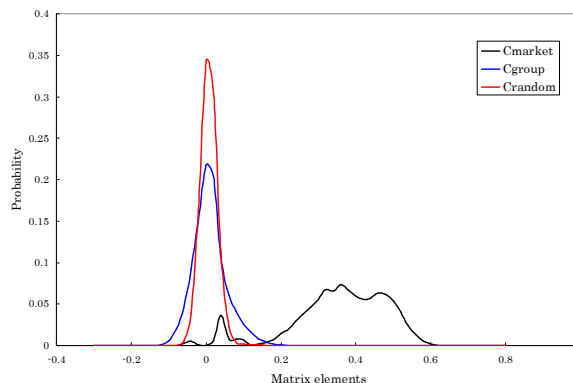


図 7: 相関行列の要素分布

これに関連して、フィルター化相関行列からポートフォリオを構築する際に、 C^{market} のみ、 C^{group} のみを使用したものと C^{mg} (前述のシミュレーションで用いたもの) を使用したものとを比較を行った。ただし、遺伝的アルゴリズムのループを 1 万回に設定したもののみである。その結果が図 8 である。

これを見ると、 C^{market} のみや、 C^{group} のみで運用結果が向上しているというわけではないことが分かる。すなわち、 C^{market} や C^{group} そのものが運用成績の向上に貢献しているというよりも、 C^{random} のノイズが除去されたことが運用成績の向上に繋がっていると考えられる。さらに言えば、 C^{mg} という、市

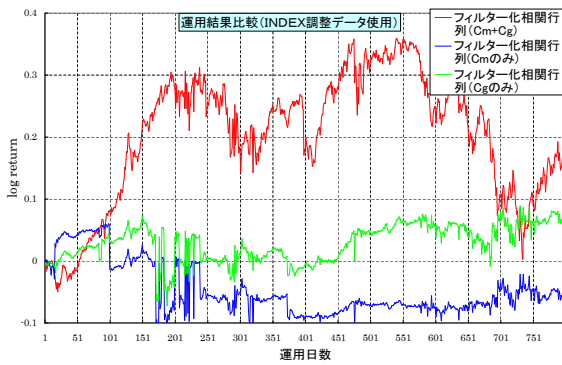


図 8: フィルター化の違いによる運用結果の比較

場本来の相関情報をもった行列にノイズによる相関情報が加わったものが従来の相関行列であるとも考えられる。

一つの仮説として、相関情報の距離の変化と平均回帰によって説明することを試みる。期間 t から Δt における相関行列 C の情報距離を $D_{\Delta t, t, C}$ とし、

$$D_{\Delta t, t, C} \equiv \sum_{j=1}^N \sum_{i=1}^N |C_{ij, t+\Delta t} - C_{ij, t}| (\forall i \neq j) \quad (10)$$

と定義する。ただし、対角成分は考慮しない。

(10) 式を基に、運用シミュレーション期間における各相関行列の情報距離を計算し、期間中の平均を求めたところ、表 3 のようになった。C に比べ C^{mg} は平

	平均情報距離
C	0.422
C^{mg}	0.390
C^{market}	0.270
C^{group}	0.268
C^{random}	0.252

表 3: 平均情報距離の比較

均情報距離が減少している。すなわち、1 期ごとの相関情報の変化が小さいということであるので、ポートフォリオ構築の最適化の程度の変化も小さいということになり、結果としてより効率のよい運用が行えると考えられる。ただし、図 8 を見れば分かる通り、平均情報距離の小さい C^{market} や C^{group} のみを用いた運用は C^{mg} を用いた運用よりも成績が悪いため、平均情報距離だけを見て運用効率を単純に測ることはできない。

平均回帰の考え方をを用いて説明すると、C はランダム要素まで含んだ相関情報であるが、ランダム要素というものは早い期間で平均回帰をする情報であるため、

ランダム要素を含んで最適化を行うと、早い周期で最適化が崩れてしまうということになる。一方、 C^{market} や C^{group} といったものは、市場の特性に基づいた情報であり、平均回帰までの期間が長いいため、最適化が崩れにくいのではないかと考えられる。

ただし、上述に関してはあくまで仮説にすぎず、フィルター化相関行列による運用成績の向上の理由の解明には更なる研究が必要である。

4 おわりに

本研究では、固有値分析を用いた相関行列のフィルタリングを行うことで、ネットワーク分析やポートフォリオ構築に関して、新たな知見が得られることを述べた。ネットワーク分析においては、通常の相関行列に比べてフィルター化相関行列は相関情報を分かりやすい形で抽出することができることがわかった。ポートフォリオ構築においては、通常の相関を利用したポートフォリオに比べ、フィルター化相関行列は効率のよい運用に貢献する可能性があることが分かった。

参考文献

- [1] V.Plerou, P.Gopikrishnan, B.Rosenow, L.A.Amaral, H.E.Stanley: Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series, *Physical Review Letters*, (1999)
- [2] L.Laloux, P.Cizeau, J-P.Bouchaud, M.Potters: Noise Dressing of Financial Correlation Matrices, *Physical Review Letters*, (1999)
- [3] J.Kwapien, S.Drozdz, J.Speth: Time scales involved in emergent market coherence, *PHYSICA A*, (2004)
- [4] Raj Kumar Pan, Sitabhra Sinha: Collective behavior of stock price movements in an emerging market, *arXiv[physics.soc-ph]*, 0704.0773v2, (2007)
- [5] <http://syrinx.q.t.u-tokyo.ac.jp/hashimoto/sociarium/>
- [6] H.Markovitz: Portfolio Selection, *Journal of Finance*, (1952)