Econophysics III: 
Financial Correlations and Portfolio Optimization

Thomas Guhr

Let’s Face Chaos through Nonlinear Dynamics, Maribor 2011
Outline

Portfolio optimization is a key issue when investing money. It is applied science and everyday work for many physicists who join the financial industry.

- importance of financial correlations
- portfolio optimization with Markowitz theory
- problem of noise dressing
- cleaning methods filtering and power mapping
- application to Swedish and US market data
- rôle of constraints in portfolio optimization
Portfolio, Risk and Correlations
Putting together a Portfolio

Portfolio 1
- ExxonMobil
- British Petrol
- Daimler
- Toyota
- ThyssenKrupp
- Voestalpine

Portfolio 2
- Sony
- British Petrol
- Daimler
- Coca Cola
- Novartis
- Voestalpine

Maribor, June/July 2011
Putting together a Portfolio

Portfolio 1
- ExxonMobil
- British Petrol
- Daimler
- Toyota
- ThyssenKrupp
- Voestalpine

Portfolio 2
- Sony
- British Petrol
- Daimler
- Coca Cola
- Novartis
- Voestalpine

correlations → diversification lowers portfolio risk!
Correlations between Stocks

visual inspection for Coca Cola and Procter & Gamble

correlations change over time!
Portfolio and Risk Management

portfolio is linear combination of stocks, options and other financial instruments

\[ V(t) = \sum_{k=1}^{K} w_k(t) S_k(t) + \sum_{l=1}^{L} w_{Cl}(t) G_{Cl}(S_l, t) + \sum_{m=1}^{M} w_{Pm}(t) G_{Pm}(S_m, t) + \ldots \]

with time–dependent weights!

portfolio or fund manager has to maximize return

- high return requires high risk: speculation
- low risk possible with hedging and diversification

find optimum for risk and return according to investors’ wishes

\[ \rightarrow \text{risk management} \]
Risk of a Portfolio

general: \( V(t) = \sum_{k=1}^{K} w_k(t) F_k(t) \) with risk elements \( F_k(t) \)

define moments within a time \( T \)

\[
\langle V(t) \rangle = \frac{1}{T} \sum_{t=1}^{T} V(t) \quad \text{and} \quad \langle V^2(t) \rangle = \frac{1}{T} \sum_{t=1}^{T} V^2(t)
\]

risk is the variance of the portfolio \( \langle V^2(t) \rangle - \langle V(t) \rangle^2 \)

often normalized \( \frac{\langle V^2(t) \rangle - \langle V(t) \rangle^2}{\langle V(t) \rangle^2} \)
Diversification — Empirically

systematic risk (market) and unsystematic risk (portfolio specific)

a wise choice of $K = 20$ stocks (or risk elements) turns out sufficient to eliminate unsystematic risk
Risk, Covariances and Correlations

if the weights \( w_k(t) \) are time–independent within the time interval \( T \) under consideration, one has

\[
\langle V^2(t) \rangle - \langle V(t) \rangle^2 = \sum_{k,l} w_k w_l \langle F_k(t) F_l(t) \rangle - \left( \sum_{k=1}^{K} w_k \langle F_k(t) \rangle \right)^2 \\
= \sum_{k,l} w_k w_l \left( \langle F_k(t) F_l(t) \rangle - \langle F_k(t) \rangle\langle F_l(t) \rangle \right) \\
= \sum_{k,l} w_k w_l \left\langle \left( F_k(t) - \langle F_k(t) \rangle \right) \left( F_l(t) - \langle F_l(t) \rangle \right) \right\rangle \\
\text{covariance matrix element } \sum_{k,l}
\]
Measuring Financial Correlations

stock prices $S_k(t), k = 1, \ldots, K$ for $K$ companies measured at times $t = 1, \ldots, T$

returns
\[
G_k(t) = \ln \frac{S_k(t + \Delta t)}{S_k(t)} \approx \frac{dS_k(t)}{S_k(t)}
\]

volatilities
\[
\sigma_k(T) = \sqrt{\langle G_k^2(t) \rangle - \langle G_k(t) \rangle^2}
\]

normalized time series
\[
M_k(t) = \frac{G_k(t) - \langle G_k(t) \rangle}{\sigma_k}
\]

correlation
\[
C_{kl}(T) = \langle M_k(t)M_l(t) \rangle = \frac{1}{T} \sum_{t=1}^{T} M_k(t)M_l(t)
\]

Maribor, June/July 2011
Financial Correlation and Covariance Matrices

correlation matrix \( C = C(T) \) is \( K \times K \) with elements \( C_{kl}(T) \)

\[
C = C(T) = \frac{1}{T}MM^\dagger
\]

covariance \( \Sigma_{kl}(T) = \left\langle \left( G_k(t) - \langle G_k(t) \rangle \right) \left( G_l(t) - \langle G_l(t) \rangle \right) \right\rangle \)

covariance and correlation are related by

\[
\Sigma_{kl}(T) = \sigma_k(T)C_{kl}(T)\sigma_l(T) , \quad \text{such that} \quad \Sigma = \sigma C \sigma ,
\]

where \( \sigma = \text{diag} \left( \sigma_1, \ldots, \sigma_K \right) \) measured volatilities
Markowitz Portfolio Optimization
Portfolio Risk and Return

**Portfolio**

\[
V(t) = \sum_{k=1}^{K} w_k S_k(t) = w \cdot S(t)
\]

\(w_k\) fraction of wealth invested, normalization

\[
1 = \sum_{k=1}^{K} w_k = w \cdot e \quad \text{with} \quad e = (1, \ldots, 1)
\]

**Desired return for portfolio**

\[
R = \sum_{k=1}^{K} w_k r_k = w \cdot r
\]

\(r_k = dS_k / S_k\) expected return for stock \(k\)

**Risk**

\[
\Omega^2 = \sum_{k,l}^{K} w_k \Sigma_{kl} w_l = w^\dagger \Sigma w = w^\dagger \sigma C \sigma w
\]
find those fractions \( w_{k}^{(\text{opt})} \) which yield the desired return \( R \) at the minimum risk \( \Omega^2 \)

Euler–Lagrange optimization problem

\[
L = \frac{1}{2} w^\dagger \Sigma w - \alpha (w \cdot r - R) - \beta (w \cdot e - 1)
\]

\( \alpha, \beta \) Langrange multipliers, \( \Sigma = \sigma C \sigma \) covariance

\[\rightarrow \quad \frac{\partial L}{\partial w} = 0, \quad \frac{\partial L}{\partial \alpha} = 0, \quad \frac{\partial L}{\partial \beta} = 0\]

system of \( K + 2 \) equations

Maribor, June/July 2011
No Constraints: Closed Form Solutions

\( K + 2 \) coupled equations read explicitly

\[
0 = \Sigma w^{(\text{opt})} - \alpha r - \beta e
\]
\[
0 = r \cdot w^{(\text{opt})} - R
\]
\[
0 = e \cdot w^{(\text{opt})} - 1
\]

if they exist, solutions are given by

\[
w^{(\text{opt})} = \frac{\Sigma^{-1} e}{e^\dagger \Sigma^{-1} e} + \frac{e^\dagger \Sigma^{-1} e R - r^\dagger \Sigma^{-1} e}{e^\dagger \Sigma^{-1} e r^\dagger \Sigma^{-1} r - (e^\dagger \Sigma^{-1} e)^2} \Sigma^{-1} \left( r - \frac{e^\dagger \Sigma^{-1} r}{e^\dagger \Sigma^{-1} e} e \right)
\]

check that they minimize risk

risk \( \Omega^2 \) is a quadratic function in desired return \( R \)
Efficient Frontier

return $R$ is a square root in the risk $\Omega^2$

inclusion of further constraints possible, for example no short selling, $w_k \geq 0$ $\rightarrow$ efficient frontier changes
Noise Dressing of Financial Correlations
Empirical Results

correlation matrices of S&P500 and TAQ data sets

eigenvalue density

distribution of spacings between the eigenvalues

true correlations are noise dressed  $\rightarrow$ DISASTER!

Laloux, Cizeau, Bouchaud, Potters, PRL 83 (1999) 1467
Plerou, Gopikrishnan, Rosenow, Amaral, Stanley, PRL 83 (1999) 1471
Quantum Chaos

result in statistical nuclear physics (Bohigas, Haq, Pandey)

universal in a huge variety of systems: nuclei, atoms, molecules, disordered systems, lattice gauge quantum chromodynamics, elasticity, electrodynamics

→ quantum “chaos” → random matrix theory

Maribor, June/July 2011
**Major Reason for the Noise**

\[
C_{kl} = \frac{1}{T} \sum_{t=1}^{T} M_k(t) M_l(t) \quad \text{we look at} \quad z_{kl} = \frac{1}{T} \sum_{t=1}^{T} a_k(t) a_l(t)
\]

with uncorrelated standard normal time series \(a_k(t)\)

\(z_{kl}\) to leading order Gaussian distributed with variance \(T\)

\[
z_{kl} = \delta_{kl} + \sqrt{1 + \delta_{kl}} \frac{1}{T} \alpha_{kl} \quad \text{with standard normal} \ \alpha_{kl}
\]

\(\rightarrow\) noise dressing \(C = C_{\text{true}} + C_{\text{random}}\) for finite \(T\)

it so happens that \(C_{\text{random}}\) is equivalent to a random matrix in the chiral orthogonal ensemble
Chiral Random Matrices

Dirac operator in relativistic quantum mechanics, \( M \) is \( K \times T \)

\[
D = \begin{bmatrix}
0 & M/\sqrt{T} \\
M^\dagger/\sqrt{T} & 0
\end{bmatrix}
\]

chiral symmetry implies off–block diagonal form

eigenvalue spectrum follows from

\[
0 = \det (\lambda 1_{K+T} - D) = \lambda^{T-K} \det (\lambda^2 1_K - MM^\dagger/T)
\]

where \( C = MM^\dagger/T \) has the form of the correlation matrix

if entries of \( M \) are Gaussian random numbers, eigenvalue density

\[
R_1(\lambda) = \frac{1}{2\pi\lambda} \sqrt{(\lambda_{\text{max}} - \lambda)(\lambda - \lambda_{\text{min}})}
\]

Maribor, June/July 2011
Correlation Matrix is Largely Random

random matrix behavior is here a DISASTER

serious doubts about practical usefulness of correlation matrices

...but: what is the meaning of the large eigenvalues?
A Model Correlation Matrix

one–factor model (called Noh’s model in physics)

\[ M_k(t) = \frac{\sqrt{p_{b(k)}} \eta_{b(k)}(t)}{\sqrt{1 + p_{b(k)}}} + \frac{\varepsilon_k(t)}{\sqrt{1 + p_{b(k)}}} \]

branch plus idiosyncratic

\( \eta_{b(k)}(t) \) and \( \varepsilon_k(t) \) are standard normal, uncorrelated time series

Maribor, June/July 2011
Explanation of the Large Eigenvalues

\[ \kappa_b \times \kappa_b \text{ matrix} \begin{bmatrix} 1 & \ldots & 1 \\ \vdots & & \vdots \\ 1 & \ldots & 1 \end{bmatrix} = ee^\dagger \quad \text{with} \quad e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \]

for \( T \to \infty \), one block has the form

\[ \frac{1}{1 + p_b} \left( p_b ee^\dagger + \kappa_b \right) \]

\( e \) itself is an eigenvector! — it yields large eigenvalue

\[ \frac{1 + p_b \kappa_b}{1 + p_b} \]

in addition, there are \( \kappa_b - 1 \) eigenvalues

\[ \frac{1}{1 + p_b} \]
Noise Reduction and Cleaning
Noise Reduction by Filtering

diagonalize $K \times K$ correlation matrix \[ C = U^{-1} \Lambda U \]

remove noisy eigenvalues \[ \Lambda = \text{diag} (\lambda_1, \ldots, \lambda_c, \lambda_{c+1}, \ldots, \lambda_K) \]

keep branch eigenvalues \[ \Lambda^{\text{(filtered)}} = \text{diag} (0, \ldots, 0, \lambda_{c+1}, \ldots, \lambda_K) \]

obtain filtered $K \times K$ correlation matrix \[ C^{\text{(filtered)}} = U^{-1} \Lambda^{\text{(filtered)}} U \]

restore normalization \[ C^{\text{(filtered)}}_{kk} = 1 \]

Parameter and Input Free Alternative?

What if “large” eigenvalue of a smaller branch lies in the bulk? Also: For smaller correlation matrices, cut–off eigenvalue $\lambda_c$ not so obvious.

There are many more noise reduction methods.

It seems that all these methods involve parameters to be chosen or other input.

Introduce the power mapping as an example for a new method. It needs little input.

The method exploits the chiral structure and the normalization.

Maribor, June/July 2011
Illustration Using the Noh Model Correlation Matrix

We look at a synthetic correlation matrix.

one–factor model (called Noh’s model in physics)

\[ M_k(t) = \frac{\sqrt{p_{b(k)} \eta_{b(k)}}(t)}{\sqrt{1 + p_{b(k)}}} + \frac{\varepsilon_k(t)}{\sqrt{1 + p_{b(k)}}} \]

branch plus idiosyncratic
Spectral Densities and Length of the Time Series

→ correlations and noise separated
Power Mapping

\[ C_{kl}(T) \quad \longrightarrow \quad \text{sign} \left( C_{kl}(T) \right) \left| C_{kl}(T) \right|^q \]

large eigenvalues (branches) only little affected

time series are effectively “prolonged”!

Heuristic Explanation

matrix element $C_{kl}$ containing true correlation $u$ and noise $v/\sqrt{T}$

$$
\left( u + \frac{v}{\sqrt{T}} \right)^q = u^q + q \frac{u^{q-1}v}{\sqrt{T}} + \mathcal{O} \left( \frac{1}{T} \right)
$$

matrix element $C_{kl}$ containing only noise $v/\sqrt{T}$

$$
\left( \frac{v}{\sqrt{T}} \right)^q = \frac{v^q}{T^{q/2}}
$$

$\rightarrow$ noise supressed for $q > 1$
optimal power $q \approx 1.5$ is automatically determined by the very definition of the correlation matrix
Internal Correlation Structure

power mapping sensitive enough to clean the internal structure
Power Mapping is a New Shrinkage Method

Shrinkage in mathematical statistics means removal of something which one does not want to be there (noise)

→ in practice: linear substraction methods
→ shrinkage parameter (and other input) needed

Power mapping is non-linear

It is parameter free and input free, because

- "chirality" 
  correlation matrix elements $C_{kl}$ are scalar products
  → noise goes like $1/\sqrt{T}$ to leading order

- normalization
  boundness $|C_{kl}| \leq 1$ → $|C_{kl}|^q \leq 1$
Sketch of Analytical Discussion

$\lambda_k(T)$ eigenvalues of $K \times K$ correlation matrix $C = C(T)$

before power mapping  

$\lambda_k(T) = \lambda_k(\infty) + \frac{v_k}{\sqrt{T}} a_k + O(1/T)$

thereafter  

$\lambda^{(q)}_k(T) = \lambda^{(q)}_k(\infty) + \frac{v^{(q)}_k}{\sqrt{T}} a^{(q)}_k + \frac{v^{(q)}_k}{Tq^{1/2}} \tilde{a}^{(q)}_k + O(1/T)$

$$
\rho_T(\lambda) = \int_{-\infty}^{+\infty} d\lambda' \frac{1}{K} \sum_{k=1}^{K} G \left( \lambda - \lambda', \frac{v^2_k}{T} \right) \rho_{\infty}(\lambda') + O(1/T)
$$

$$
\rho^{(q)}_T(\lambda) = \int_{-\infty}^{+\infty} d\lambda' \frac{1}{K} \sum_{k=1}^{K} G \left( \lambda - \lambda', \frac{(\tilde{v}^{(q)}_k)^2}{Tq} \right) \rho_T(\lambda') \bigg|_{v^{(q)}_k} + O(1/T)
$$

Maribor, June/July 2011
Result for Power–Mapped Noh Model

\[ M_k(t) = \frac{\sqrt{p_{b(k)} \eta_{b(k)}}(t)}{\sqrt{1 + p_{b(k)}}} + \frac{\varepsilon_k(t)}{\sqrt{1 + p_{b(k)}}} \]

\(B\) branches, sizes \(\kappa_b, b = 1, \ldots, B\)
\(\kappa\) companies in no branch

\[ \rho_T^{(q)}(\lambda) = (K - \kappa - B)G \left( \lambda - \mu_B^{(q)}, \frac{(v_B^{(q)})^2}{T} \right) + \kappa G \left( \lambda - 1, \frac{(v_0^{(q)})^2}{Tq} \right) \]

\[ + \sum_{b=1}^B \delta \left( \lambda - \left( 1 + (\kappa_b - 1) \left( \frac{p_b}{1 + p_b} \right)^q \right) \right) \]

where \(\mu_B^{(q)} = 1 - \frac{1}{B} \sum_{b=1}^B \left( \frac{p_b}{1 + p_b} \right)^q\)
Application to Market Data
Markowitz Optimization after Noise Reduction

**portfolio optimization**

**Markowitz theory**

Swedish stock returns
197 companies, daily, from July 12, 1999 to July 18, 2003
sample: one year — evaluation steps: one week

Standard & Poor’s 500
100 most actively traded stocks, daily data 2002 to 2006
sample: 150 days — evaluation steps: 14 days

Maribor, June/July 2011
**Swedish Stocks — No Constraints**

Markowitz theory with desired return of 0.3% per week

<table>
<thead>
<tr>
<th></th>
<th>yearly actual risk [%]</th>
<th>yearly actual return [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample</td>
<td>20.7</td>
<td>11.1</td>
</tr>
<tr>
<td>power mapped</td>
<td>11.3</td>
<td>5.0</td>
</tr>
<tr>
<td>filtered</td>
<td>11.4</td>
<td>10.5</td>
</tr>
</tbody>
</table>

Maribor, June/July 2011
Swedish Stocks — Constraint: No Short Selling

Markowitz theory with desired return of 0.1% per week

<table>
<thead>
<tr>
<th></th>
<th>yearly actual risk [%]</th>
<th>yearly actual return [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample</td>
<td>10.1</td>
<td>0.5</td>
</tr>
<tr>
<td>power mapped</td>
<td>9.9</td>
<td>1.1</td>
</tr>
<tr>
<td>filtered</td>
<td>9.9</td>
<td>0.7</td>
</tr>
</tbody>
</table>
Swedish Stocks — Weights

no constraints

constraint: no short selling

→ less rigid: filtering seems favored

more rigid: power mapping seems favored

Maribor, June/July 2011
Mean Value of Correlation Matrix

in a sampling period

\[ c = \frac{1}{K^2} \sum_{k,l} C_{kl} \]

very similar curves!

important: Markowitz optimization is invariant under scaling

\[ C \rightarrow \gamma C \quad \text{for all} \quad \gamma > 0 \]

power mapped \( C' \) can be readjusted with \( \frac{c^{(\text{original})}}{c^{(\text{power mapped})}} \)
Standard & Poor’s — Adjusted Power

\[ K = 100, \ T = 150, \ q_{\text{opt}} = 1.8 \]

Mean realized risk
- \( C^{\text{sample}} \) \quad 1.548e-2
- \( C^{\text{filter}} \) \quad 0.809e-2
- \( C^{(q)} \) \quad 0.812e-2

Mean realized return
- \( C^{\text{sample}} \) \quad 0.42e-4
- \( C^{\text{filter}} \) \quad 5.45e-4
- \( C^{(q)} \) \quad 5.90e-4

constraint: no short selling
$K = 100, \ T = 150, \ q_{\text{opt}} = 1.8$

constraint: no short selling

Power-mapping yields good risk-reduction for wide range of $q$ values
Summary and Conclusions

- Portfolio risk depends on correlations
- Markowitz optimization is an Euler–Lagrange problem
- Correlations are noise dressed
- Two noise reduction methods discussed: filtering and power mapping
- Both are good at reducing risk, perform differently in the presence of constraints (no short selling)
- Example for everyday work in financial industry