Cleaning Correlation Matrices from Random Noise

Gábor Papp, Szilárd Pafka, Maciej A. Nowak, Imre Kondor

Eötvös University, Budapest,
Jagellonian University, Cracow,
Collegium Budapest, Budapest,

Morgan Stanley
Quantitative and Mathematical Finance Conference
October 20-21, 2005, Budapest, Hungary

Portfolio problem

- How to construct the "optimal" portfolio out of N, possibly correlated instruments?
  - optimal return?
  - optimal risk (VAR)?
- H. Markowitz, J. Fin. **7** 91 (1952)
- It is important to know the behaviour of the instruments in the past: data collection.
- Past returns do not tell us the expected return (autocorrealtion $\sim$ 15m.
- Past volatilities are rather stable and have predictive power to the future values ("adiabatic process").

Portfolio problem

- How to construct the "optimal" portfolio out of N, possibly correlated instruments?
  - optimal return?
  - optimal risk (VAR)?
- H. Markowitz, J. Fin. **7** 91 (1952)
- It is important to know the behaviour of the instruments in the past: data collection.
- Past returns do not tell us the expected return (autocorrealtion $\sim$ 15m.
- Past volatilities are rather stable and have predictive power to the future values ("adiabatic process").

Covariance matrix

- The basic object of the portfolio analysis is the Covariance Matrix.

$$C_{ij} = \langle x_i(t) x_j(t) \rangle_t = \frac{1}{T} \sum_{t=1}^{T} x_i(t) x_j(t).$$

with price changes

$$x(t) = \log \frac{P(t)}{P(t-1)} - Div(t).$$

- Goal: minimize risk of the portfolio (with a gaussian assumption)

$$R = \langle p_i C_{ij} p_j \rangle$$

with portfolio entries p_i's.

Covariance matrix

- The basic object of the portfolio analysis is the Covariance Matrix.

$$C_{ij} = \langle x_i(t) x_j(t) \rangle_t = \frac{1}{T} \sum_{t=1}^{T} x_i(t) x_j(t).$$

with price changes

$$x(t) = \log \frac{P(t)}{P(t-1)} - Div(t).$$

- Goal: minimize risk of the portfolio (with a gaussian assumption)

$$R = \langle p_i C_{ij} p_j \rangle$$

with portfolio entries p_i's.

Markowitz solution

- Markowitz's solution to the problem (without considering return information)

$$p_i = \frac{\sum_k C_{ik}^{-1}}{\sum_{i,k} C_{ik}^{-1}} \qquad \text{(Markowitz)}$$

with a minimal risk

$$R_{min} = \frac{1}{\sum_{i,j} C_{ij}^{-1}}$$

- Inclusion of returns to the minimization procedure increases risk.

- In case of non-Gaussian processes the form (definition) of the $\mathbf{C}$ covariance matrix is extremely important, and has an effect on the optimization procedure.

Markowitz solution

- Markowitz's solution to the problem (without considering return information)

$$p_i = \frac{\sum_k C^{-1}_{ik}}{\sum_{i,k} C^{-1}_{ik}} \qquad \text{(Markowitz)}$$

with a minimal risk

$$R_{\min} = \frac{1}{\sum_{i,j} C^{-1}_{ij}}$$

- Inclusion of returns to the minimization procedure increases risk.
- In case of non-Gaussian processes the form (definition) of the $\mathbf{C}$ covariance matrix is extremely important, and has an effect on the optimization procedure.

Noise

- The price-change series matrix typically is of size $N \times T$, timeseries of length T for N instruments,

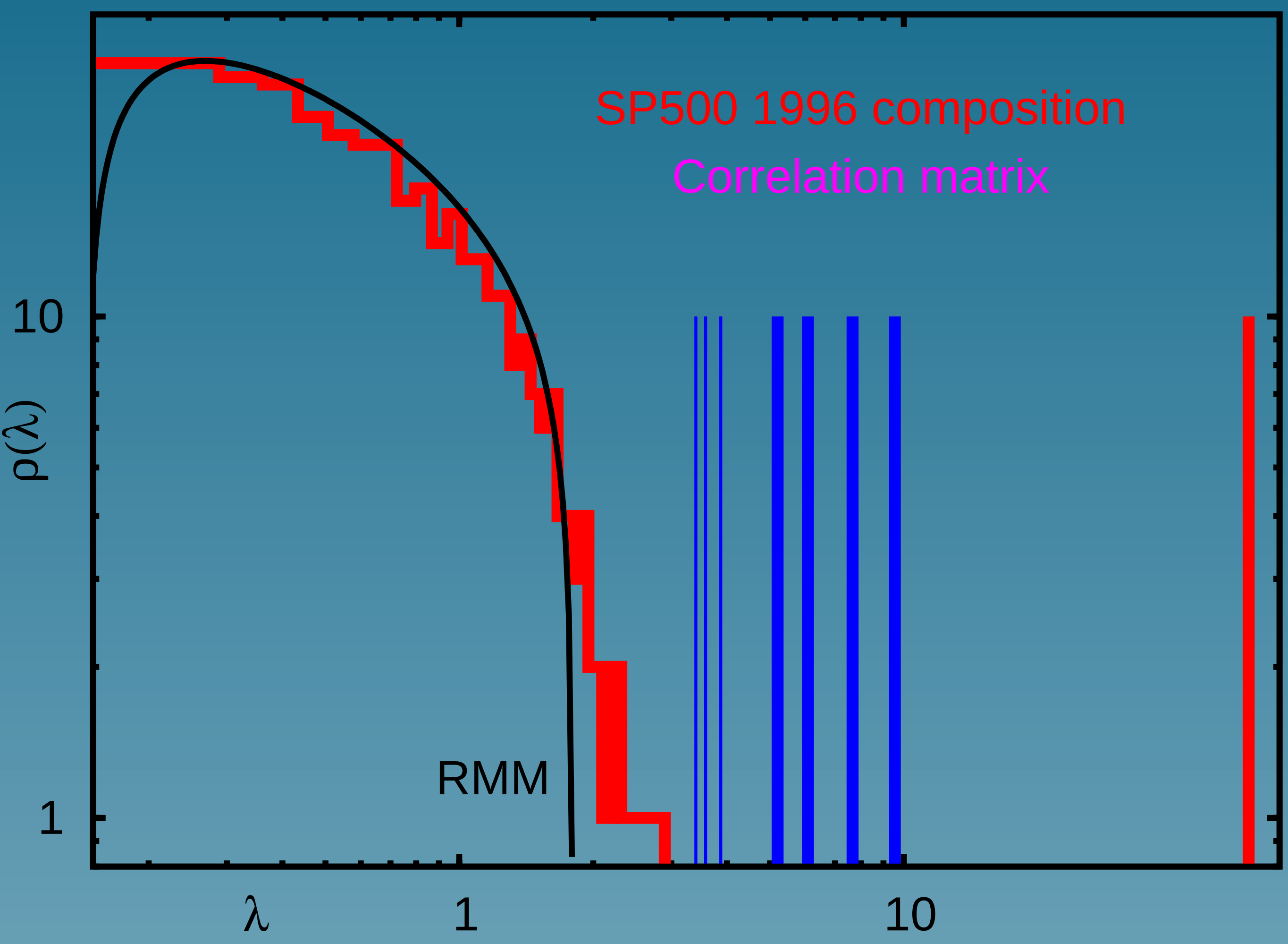
$$\mathbf{A} = \underbrace{\begin{pmatrix} x_1(1) & x_1(2) & \ldots & x_1(T) \\ x_2(1) & x_2(2) & \ldots & x_2(T) \\ \vdots & & \ddots & \\ x_N(1) & x_N(2) & \ldots & x_N(T) \end{pmatrix}}_{T} \left.\vphantom{\begin{pmatrix} x \\ x \\ x \\ x \end{pmatrix}}\right\} N$$

- The Covariance matrix is $\mathbf{C} = \frac{1}{T}\mathbf{A}\mathbf{A}^T$.

- Entries $x_i(t)$ are usually correlated, and

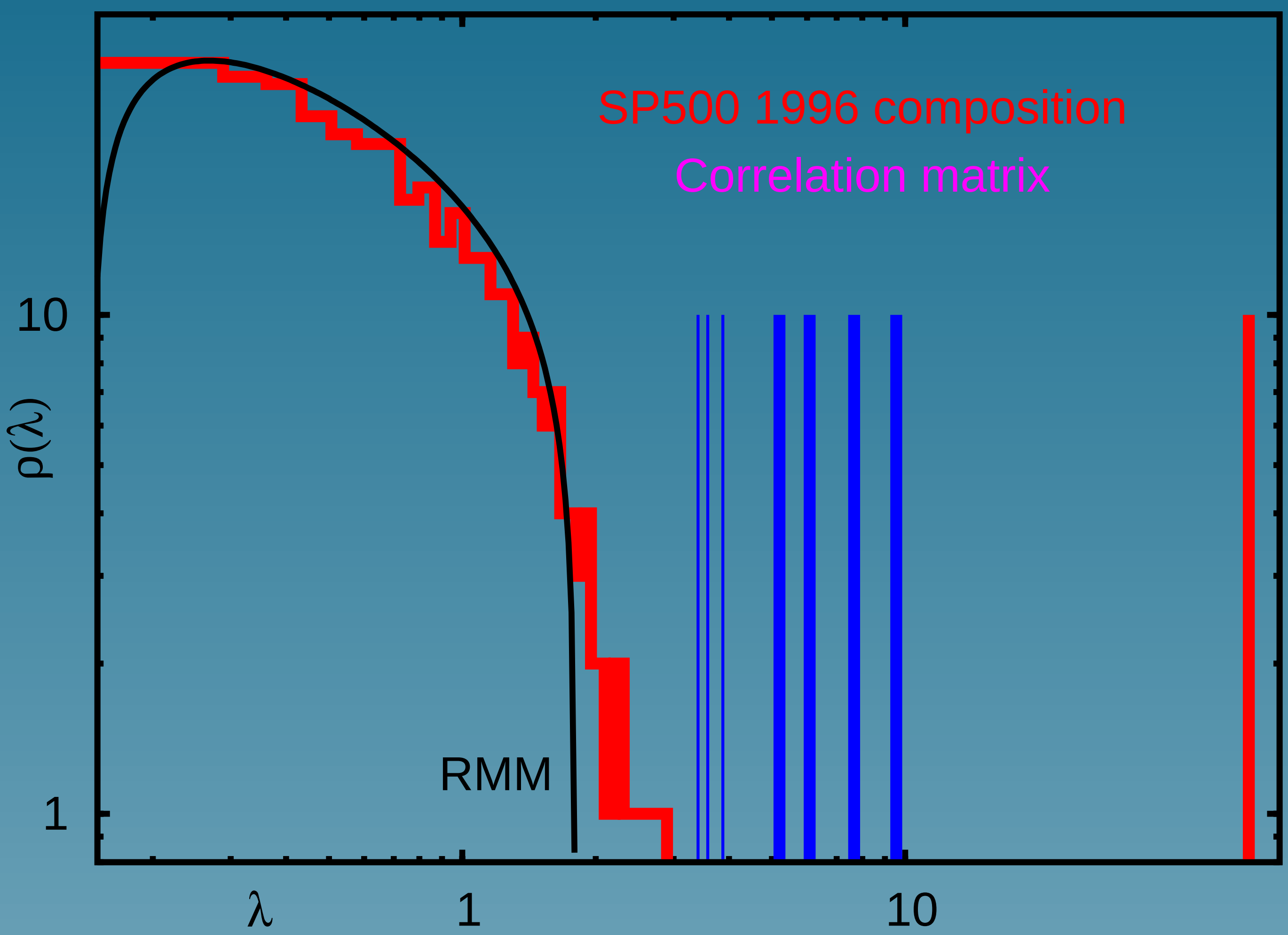
- are dressed with NOISE, hence the Covariance matrix is also NOISY.

Spectrum of the Correlation Matrix



RMM sectors market

Spectrum of the Correlation Matrix



RMM sectors market

Wishart matrices

- $\mathbf{M}M^\dagger/T$, where $\mathbf{M}$ is an $N \times T$ matrix
- Wishart, Biometrika A**20** (1928) 32.
- Marchenko, Pastur, Math. USSR-Sbornik, **1** (1967) 457.
- spectrum (with $r = N/T$, $\mathbf{M}$ being random gaussian)

$$\varrho(\lambda) = \frac{1}{2r\pi\lambda}\sqrt{(\lambda - \lambda_{min})(\lambda_{max} - \lambda)},$$

$$\lambda_{min,max} = (1 \pm \sqrt{r})^2$$

- if $r > 1$ there are additional zero modes ($r\delta_0$).
- for $N/T \to 0$ only the "real" correlation survives at $\lambda = \mathbf{1}$.

Wishart matrices

- $\mathbf{M}M^\dagger/T$, where $\mathbf{M}$ is an $N \times T$ matrix
- Wishart, Biometrika A**20** (1928) 32.
- Marchenko, Pastur, Math. USSR-Sbornik, **1** (1967) 457.
- spectrum (with $r = N/T$, $\mathbf{M}$ being random gaussian)

$$\varrho(\lambda) = \frac{1}{2r\pi\lambda}\sqrt{(\lambda - \lambda_{min})(\lambda_{max} - \lambda)},$$

$$\lambda_{min,max} = (1 \pm \sqrt{r})^2$$

- if $r > 1$ there are additional zero modes $(r\delta_0)$.
- for $N/T \to 0$ only the “real” correlation survives at $\lambda = \mathbf{1}$.

How to deal with Random Matrices?

- D.V. Voiculescu, *Invent. Math.* **104** (1991) 201.
- Free Random Variables
- Eg.: Gaussian distribution (numbers)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{x^2}{2\sigma^2}}$$

- What is the distribution for $x_1 + x_2$, both Gaussian?

$$p_2(S = x_1 + x_2) = p(x_1) \otimes p(x_2), \quad \text{or}$$

$$\varphi_2(q) = \varphi(q)\varphi(q), \quad \text{with} \qquad \varphi(q) = \int dx\, e^{iqx}\; p(x), \text{or}$$

$$\ln \varphi_2(q) = \ln \varphi(q) + \ln \varphi(q)$$ ADDITION LAW

How to deal with Random Matrices?

- D.V. Voiculescu, *Invent. Math.* **104** (1991) 201.
- Free Random Variables
- Eg.: Gaussian distribution (numbers)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$$

- What is the distribution for $x_1 + x_2$, both Gaussian?

$$p_2(S = x_1 + x_2) = p(x_1) \otimes p(x_2), \quad \text{or}$$

$$\varphi_2(q) = \varphi(q)\varphi(q), \quad \text{with} \qquad \varphi(q) = \int dx\, e^{iqx}\, p(x), \text{or}$$

$$\ln \varphi_2(q) = \ln \varphi(q) + \ln \varphi(q) \quad \text{ADDITION LAW}$$

How to deal with Random Matrices?

- D.V. Voiculescu, *Invent. Math.* **104** (1991) 201.
- Free Random Variables
- Eg.: Gaussian distribution (numbers)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$$

- What is the distribution for $x_1 + x_2$, both Gaussian?

$$p_2(S = x_1 + x_2) = p(x_1) \otimes p(x_2), \quad \text{or}$$

$$\varphi_2(q) = \varphi(q)\varphi(q), \quad \text{with} \qquad \varphi(q) = \int dx\, e^{iqx}\, p(x), \text{or}$$

$$\ln \varphi_2(q) = \ln \varphi(q) + \ln \varphi(q)$$ ADDITION LAW

How to deal with Random Matrices?

- D.V. Voiculescu, *Invent. Math.* **104** (1991) 201.
- Free Random Variables
- Eg.: Gaussian distribution (numbers)

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

- What is the distribution for $x_1 + x_2$, both Gaussian?

$$p_2(S = x_1 + x_2) = p(x_1) \otimes p(x_2), \quad \text{or}$$

$$\varphi_2(q) = \varphi(q)\varphi(q), \quad \text{with} \qquad \varphi(q) = \int dx\, e^{iqx}\, p(x), \text{or}$$

$$\ln \varphi_2(q) = \ln \varphi(q) + \ln \varphi(q) \quad \text{ADDITION LAW}$$

How to deal with Random Matrices?

- D.V. Voiculescu, *Invent. Math.* **104** (1991) 201.
- Free Random Variables
- Eg.: Gaussian distribution (numbers)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{x^2}{2\sigma^2}}$$

- What is the distribution for $x_1 + x_2$, both Gaussian?

$$p_2(S = x_1 + x_2) = p(x_1) \otimes p(x_2), \quad \text{or}$$

$$\varphi_2(q) = \varphi(q)\varphi(q), \quad \text{with} \qquad \varphi(q) = \int dx\, e^{iqx}\; p(x), \text{or}$$

$$\ln \varphi_2(q) = \ln \varphi(q) + \ln \varphi(q) \quad \text{ADDITION LAW}$$

How to deal with Random Matrices?

- Do we have similar for MATRICES?

- M is an $N \times N$ random Gaussian matrix:

  Resolvent $G(z) = \int dM\, e^{-N\,V(M)}\, \frac{1}{N}\,\mathrm{Tr}_N\, \frac{1}{z-M}$,

  with $V(M) = \frac{1}{2}\,\mathrm{Tr}\; M^2$.

- for Gaussian matrices it fulfilles a simple algebraic eq:

$$G(z) = (z - G(z))^{-1} \quad \rightarrow \quad G(z) = \frac{1}{2}\left(z - \sqrt{z^2 - 4}\right)$$

and spectral density $\varrho(\lambda) = -\frac{1}{\pi}\lim_{\epsilon\to 0}\,\mathrm{Im}\; G(z)\big|_{z=\lambda+i\epsilon}$,

$$\varrho(\lambda) = \frac{1}{2\pi}\sqrt{4-\lambda^2}$$

Wigner's semicircle law.

How to deal with Random Matrices?

- Do we have similar for MATRICES?
- M is an $N \times N$ random Gaussian matrix:

  Resolvent $G(z) = \int dM \, e^{-N\,V(M)} \frac{1}{N} \mathrm{Tr}_N \frac{1}{z-M}$,

  with $V(M) = \frac{1}{2}$ Tr M^2.

- for Gaussian matrices it fulfilles a simple algebraic eq:

$$G(z) = (z - G(z))^{-1} \quad \rightarrow \quad G(z) = \frac{1}{2}\left(z - \sqrt{z^2 - 4}\right)$$

  and spectral density $\varrho(\lambda) = -\frac{1}{\pi} \lim_{\epsilon \to 0} \mathrm{Im}\, G(z)\big|_{z=\lambda+i\epsilon}$,

$$\varrho(\lambda) = \frac{1}{2\pi}\sqrt{4 - \lambda^2}$$

  Wigner's semicircle law.

How to deal with Random Matrices?

- Do we have similar for MATRICES?

- M is an $N \times N$ random Gaussian matrix:

$$\text{Resolvent } G(z) = \int dM\, e^{-N\,V(M)}\, \frac{1}{N}\, \text{Tr}_N\, \frac{1}{z-M},$$

with $V(M) = \frac{1}{2}\, \text{Tr}\, M^2$.

- for Gaussian matrices it fulfilles a simple algebraic eq:

$$G(z) = (z - G(z))^{-1} \quad \rightarrow \quad G(z) = \frac{1}{2}\left(z - \sqrt{z^2-4}\right)$$

and spectral density $\varrho(\lambda) = -\frac{1}{\pi} \lim_{\epsilon\to 0} \text{Im}\, G(z)|_{z=\lambda+i\epsilon}$,

$$\varrho(\lambda) = \frac{1}{2\pi}\sqrt{4-\lambda^2}$$

Wigner's semicircle law.

Addition Law

- Condsider random matrices M_1 and M_2
- $[M_1, M_2] \neq 0!$ do not commute
- What is the spectrum of $M_1 + M_2$ on ensemle average?

$$G_{1+2}(z) = \int dM_1 dM_2 \, e^{-NV(M_1)} \, e^{-NV(M_2)} \frac{1}{N} \mathrm{Tr}_N \frac{1}{z - M_1 - M_2}$$

- Define $R(z)$ as $G[R(z) + 1/z] = z$, and R transform is ADDITIVE for matrix addition:

$$R_{1+2}(z) = R_1(z) + R_2(z)$$

- Eg.: add two matrices from the same Gaussian distribution:

$$R_1(z) = R_2(z) = z \to R_{1+2}(z) = 2z \to G(z) = \frac{1}{4}\left(z - \sqrt{z^2 - 8}\right)$$

Addition Law

- Condsider random matrices M_1 and M_2
- $[M_1, M_2] \neq 0!$ do not commute
- What is the spectrum of $M_1 + M_2$ on ensemle average?

$$G_{1+2}(z) = \int dM_1 dM_2 \, e^{-N\,V(M_1)} \, e^{-N\,V(M_2)} \, \frac{1}{N} \mathrm{Tr}_N \, \frac{1}{z - M_1 - M_2}$$

- Define $R(z)$ as $G[R(z) + 1/z] = z$, and R transform is ADDITIVE for matrix addition:

$$R_{1+2}(z) = R_1(z) + R_2(z)$$

- Eg.: add two matrices from the same Gaussian distribution:

$$R_1(z) = R_2(z) = z \rightarrow R_{1+2}(z) = 2z \rightarrow G(z) = \frac{1}{4}\left(z - \sqrt{z^2 - 8}\right)$$

Addition Law

- Condsider random matrices M_1 and M_2
- $[M_1, M_2] \neq 0!$ do not commute
- What is the spectrum of $M_1 + M_2$ on ensemle average?

$$G_{1+2}(z) = \int dM_1 dM_2 \, e^{-NV(M_1)} \, e^{-NV(M_2)} \, \frac{1}{N} \mathrm{Tr}_N \, \frac{1}{z - M_1 - M_2}$$

- Define $R(z)$ as $G[R(z) + 1/z] = z$, and R transform is ADDITIVE for matrix addition:

$$R_{1+2}(z) = R_1(z) + R_2(z)$$

- Eg.: add two matrices from the same Gaussian distribution:

$$R_1(z) = R_2(z) = z \to R_{1+2}(z) = 2z \to G(z) = \frac{1}{4}\left(z - \sqrt{z^2 - 8}\right)$$

Addition Law

- Condsider random matrices M_1 and M_2
- $[M_1, M_2] \neq 0!$ do not commute
- What is the spectrum of $M_1 + M_2$ on ensemle average?

$$G_{1+2}(z) = \int dM_1 dM_2 \, e^{-NV(M_1)} \, e^{-NV(M_2)} \frac{1}{N} \mathrm{Tr}_N \frac{1}{z - M_1 - M_2}$$

- Define $R(z)$ as $G[R(z) + 1/z] = z$, and R transform is ADDITIVE for matrix addition:

$$R_{1+2}(z) = R_1(z) + R_2(z)$$

- Eg.: add two matrices from the same Gaussian distribution:

$$R_1(z) = R_2(z) = z \to R_{1+2}(z) = 2z \to G(z) = \frac{1}{4}\left(z - \sqrt{z^2 - 8}\right)$$

and the resulting distribution is the rescaling by $\sqrt{2}$:

$$\varrho(\lambda) = \frac{1}{4\pi}\sqrt{8-\lambda^2}$$

Multiplication Law

- What is the spectrum of $M_1 \cdot M_2$ on ensemle average? (positive definite, symmetric matrices)
- Define $S(z)$ as

$$S(z) = \frac{1+z}{z}\chi(z), \qquad \frac{1}{\chi(z)}G[\frac{1}{\chi(z)}] - 1 = M[\frac{1}{\chi(z)}] = z$$

where M is the generating function.

- $S(z)$ is MULTIPLICATIVE

$$S_{1\cdot 2} = S_1(z) \cdot S_2(z)$$

Multiplication Law: Example

- consider an $N \times T$ matrix $\mathbf{X}$, and construct $\frac{1}{T}\mathbf{X}\mathbf{X}^{\mathbf{T}}$ (symmetric, positive definite matrix)

- Spectrum for the product of $\mathbf{Y}$ $N \times N$ ($N >$T) matrices is known:

$$\varrho(z = \lambda^2) = \frac{\varrho_G(\lambda)}{\lambda} = \frac{1}{2\pi}\frac{\sqrt{4-z}}{\sqrt{z}}$$

- To get the proper spectrum of $\frac{1}{T}\mathbf{X}\mathbf{X}^{\mathbf{T}}$, $\frac{1}{T}\mathbf{Y}\mathbf{Y}^{\mathbf{T}}$ should be multiplied by the projector $\mathbf{P} = diag(1, ..., 1, 0, ...0)$:

$$S_{YY^T} = \frac{1}{1+z}, \qquad S_P = \frac{1+z}{m+z}$$

with $m = T/N$. Finally, with the inverse transformation one recovers

$$\varrho(\lambda) = (1-m)\delta(\lambda)\, \tfrac{1}{2\pi}\sqrt{(\lambda_> - \lambda)(\lambda - \lambda_<)}$$

with $\lambda_{<>} = (1 \pm \sqrt{m})^2$ (Marchenko-Pastur result).

Multiplication Law: Finances

- Let $\mathbf{C}$ be the spatial (instrument) correlation, $\mathbf{A}$ the temporal correlation. Constructing the *empirical* correlation matrix

$$G(z) = \frac{1}{N}\left\langle \mathrm{Tr}_N \frac{1}{z - \frac{1}{T}\mathbf{X}\mathbf{X}^T} \right\rangle \quad \text{with measure} \quad \int d\mathbf{X}\, e^{-\mathrm{Tr}_T \mathbf{X}^T \mathbf{A}^{-1} \mathbf{X}}$$

 one arrives at

$$M_{\mathbf{A}}(w) = m(z) \quad \text{with} \qquad w = \frac{z}{rm(z) M_{\mathbf{C}}^{-1}[rm(z)]}$$

 with $m(z)$ the generating function of the *empirical* matrix and $M_{\mathbf{A}}(z)$, $M_{\mathrm{C}}(z)$ the generating functions of true correlation matrices $\mathbf{A}$ and $\mathbf{C}$, respectively.

- The problem is INVERTIBLE! (Cleaning – Dressing)

Risks

- Clean system (Markowitz):

covariance matrix: $\mathbf{C}^{(0)}$,

portfolio entries: $$p_i^{(0)} = \frac{\sum_k {C^{(0)}}^{-1}_{ik}}{\sum_{i,k} {C^{(0)}_{ik}}^{-1}}.$$

- Noisy system:

covariance matrix: $\mathbf{C}$,

portfolio entries: $$p_i = \frac{\sum_k C^{-1}_{ik}}{\sum_{i,k} C^{-1}_{ik}}.$$

- Real risk ratio:

$$q_0^2 = \frac{\langle p|\mathbf{C}^{(0)}|p\rangle}{\langle p^{(0)}|\mathbf{C}^{(0)}|p^{(0)}\rangle}$$

Risks II

- This ratio can be computed analytically from Random Matrix Theory (gaussian case):

$$q_0^2 = \frac{\int d\lambda \, \frac{1}{\lambda^2} \, \varrho(\lambda)}{\left(\int d\lambda \, \frac{1}{\lambda} \, \varrho(\lambda) \right)^2} \quad = \quad \frac{1}{1-r}$$

- For long enough time series r is small, and the optimal portfolio is fairly reconstructed from the noisy, empirical covariance matrix.

- For a large portfolio r is close to 1, and the optimization is completely meaningless!

- For $r > 1$ the system is completely unpredictable (the number of unknown parameters is larger than the number of measurements!)

Risks II

- This ratio can be computed analytically from Random Matrix Theory (gaussian case):

$$q_0^2 = \frac{\displaystyle\int d\lambda \, \frac{1}{\lambda^2} \, \varrho(\lambda)}{\left(\displaystyle\int d\lambda \, \frac{1}{\lambda} \, \varrho(\lambda) \right)^2} \quad = \quad \frac{1}{1-r}$$

- For long enough time series r is small, and the optimal portfolio is fairly reconstructed from the noisy, empirical covariance matrix.
- For a large portfolio r is close to 1, and the optimization is completely meaningless!
- For $r > 1$ the system is completely unpredictable (the number of unknown parameters is larger than the number of measurements!)

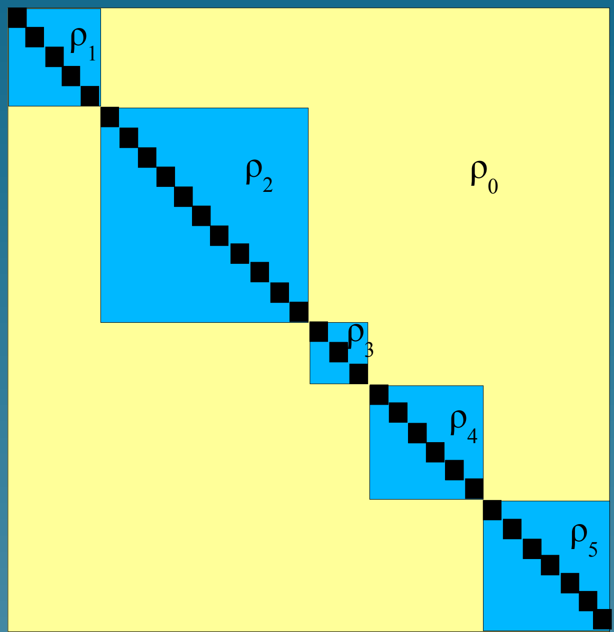
What can be done?

- In case we have some additional information about the covariance matrix, actually, the number of measurements may be less, than the number of elements in the covariance matrix, and still a reasonable optimization may be performed:

- Analysis shows, that typically there are a couple of large eigenvalues, and the rest is very close to the Random Matrix Model.

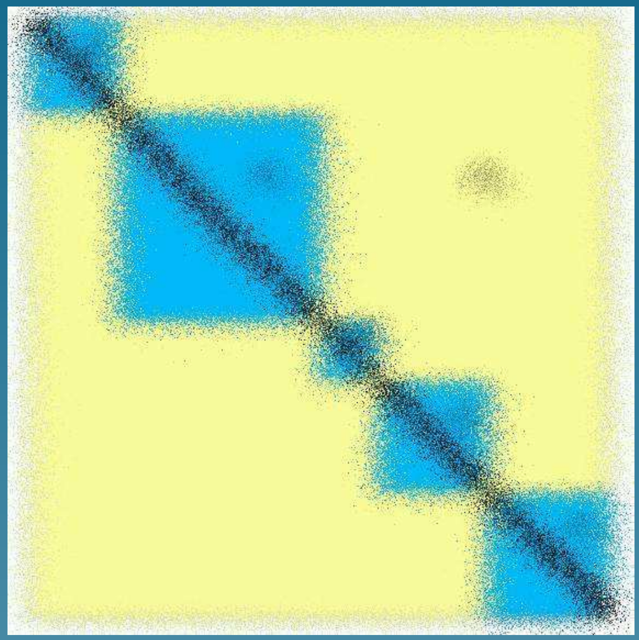
- The large eigenvalues can be associated with "sectors".
  - keep the entries associated with the sectors, and substitute the low lying part with the average eigenvalue.
  - use "cleaning" on the empirical covariance matrix assuming a certain number of sectors.

Sectorization



$\longrightarrow$

(Gaussian) noise



- Coupling between timeseries (Noh, Phys. Rev. **E61** (2000) 5981):
  - ϱ_0 : "market"
  - ϱ_i : sectors (energy, telecommunication, etc.)
  - There are M different degenerate, and M large eigenvalues for a matrix with M sectors.

- Noise dims the real structure.

Cleaning: How does it work on SP500?

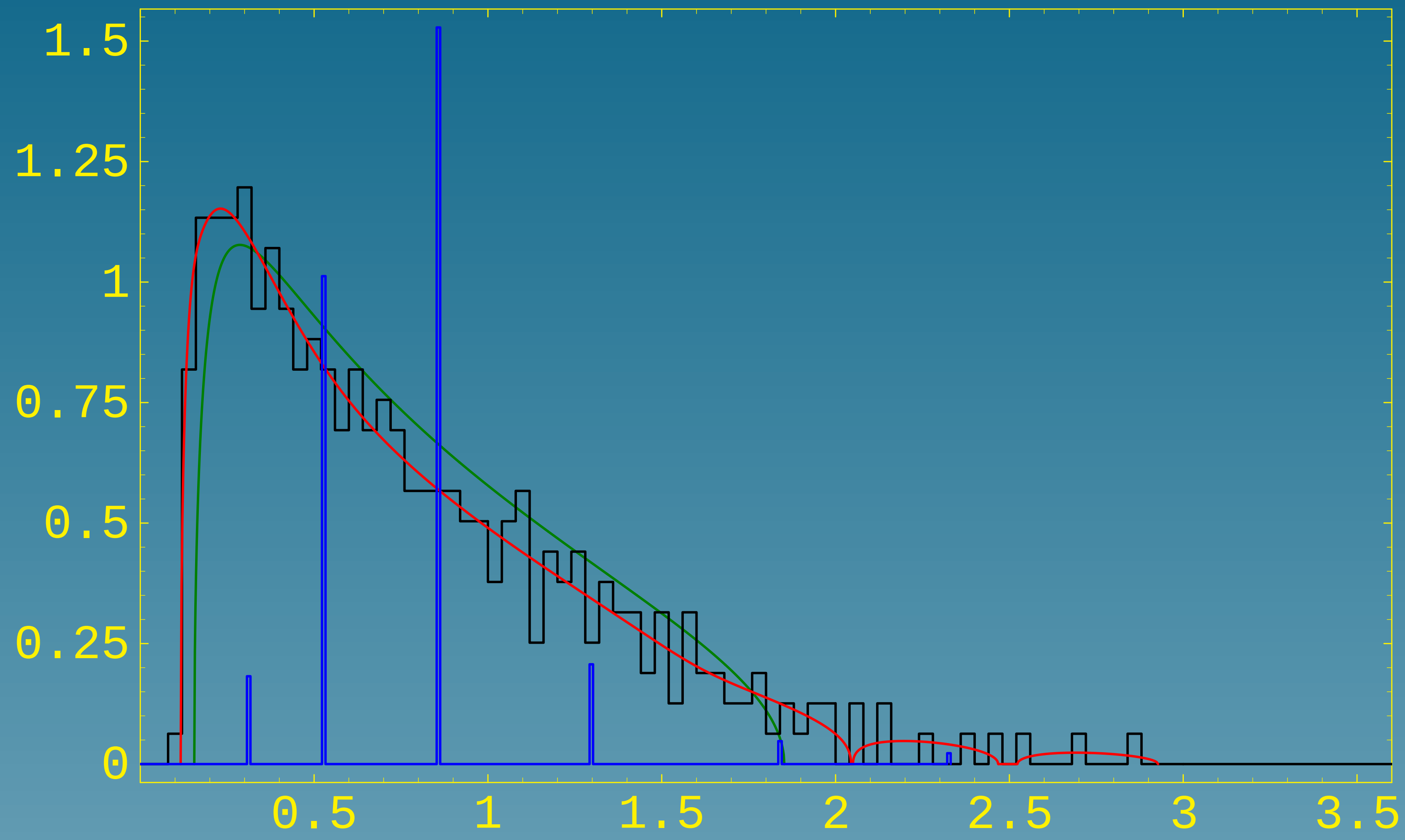
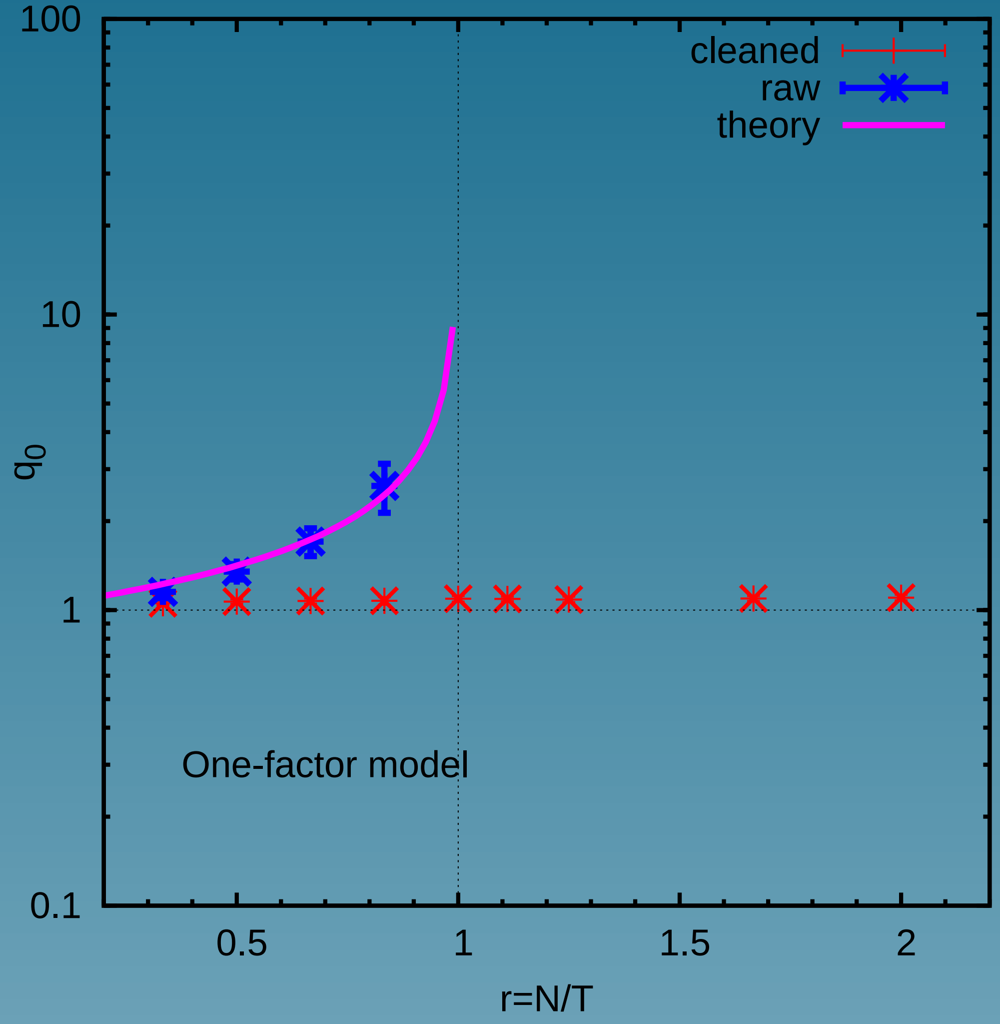


Figure 1: Approximation of the SP500 covariance matrix with a 6 “sector” assumption.

Benchmarking

- Check, how well the method works:
- Start with a know clean covariance matrix (assuming be the same as the correlation matrix).
- Add gaussian noise (generate finite price-change timeseries).
- reconstruct the best estimate of the clean covariance matrix using the cleaning technique.
- Compare the risk of the cleaned, empirical portfolio to the risk of the clean portfolio, using Markowitz (q_0).

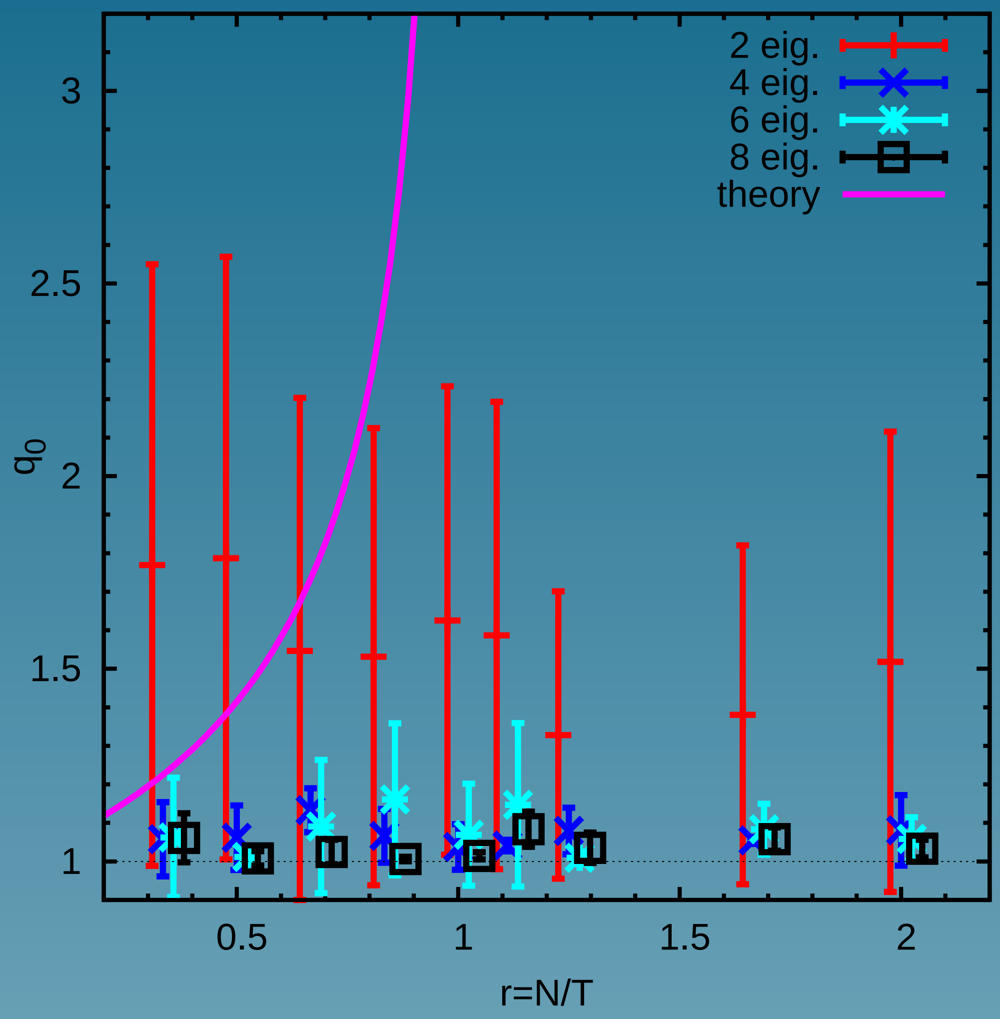
One factor model



Test of one-factor model for different values of r with fixed $N = 100$ and different number of sectors searched.

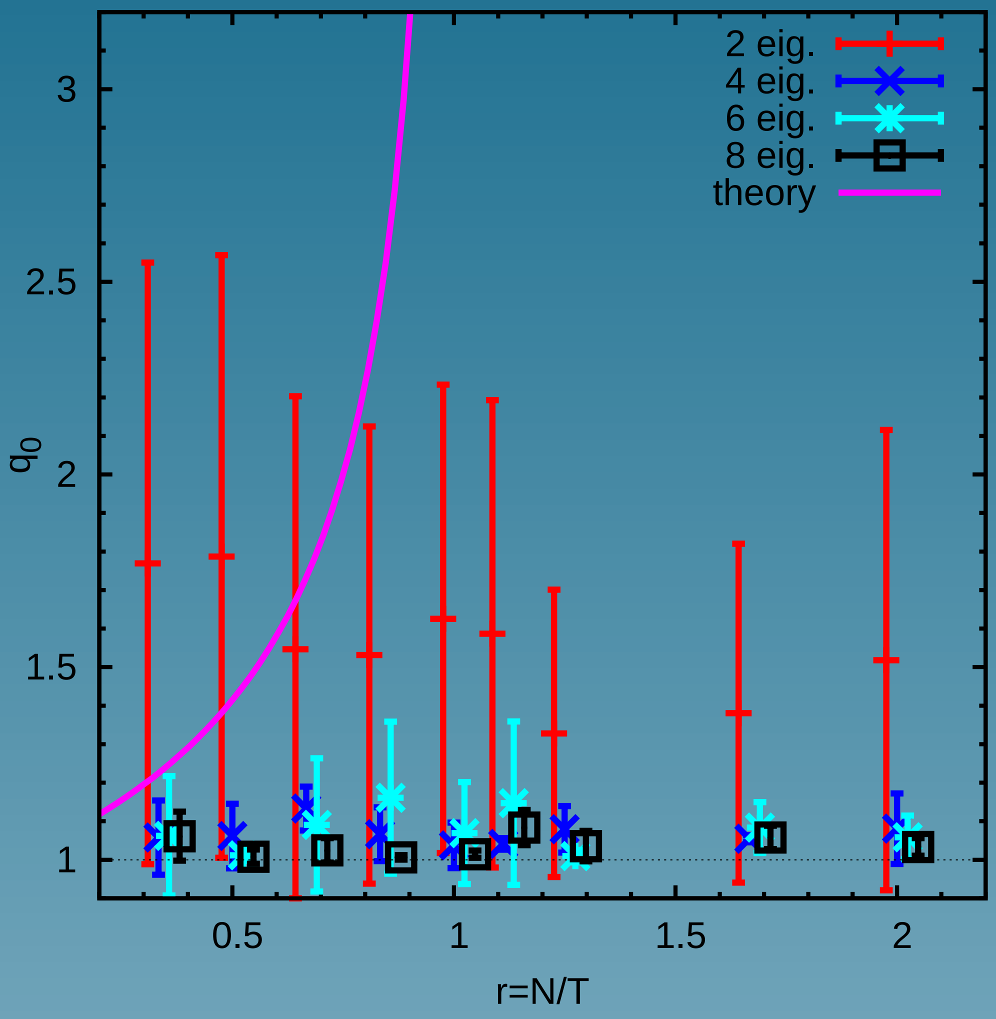
- perfect reconstruction
- does not depend on the number of sectors searched

Market-plus-sectors



Test of market-plus-sector model for different values of r with fixed $N = 100$ and different number of eigenvalues searched (2, 4, 6 and 8 eigenvalues).

Market plus-sectors



- searching for at least 4 eigenvalues gives a very good reconstruction
- no pronounced dependence on the number of eigenvalues searched
- Using less eigenvalues than necessary, worsens the situation: even the raw empirical matrix provides a better solution!

Summary

- Finite value of the ration r=N/T causes a huge problem in portfolio optimization.
- Cleaning is possible using a Random Matrix Model.
- Cleaning may be used to optimize a portfolio.
- The method is robust, i.e. in case the number of sectors is unknown, overestimating it does not cause any problem.
- Cleaning normally gives better results, than a simple RMM based spectrum decompositon.

Summary

- Finite value of the ration $r = N/T$ causes a huge problem in portfolio optimization.
- Cleaning is possible using a Random Matrix Model.
- Cleaning may be used to optimize a portfolio.
- The method is robust, i.e. in case the number of sectors is unknown, overestimating it does not cause any problem.
- Cleaning normally gives better results, than a simple RMM based spectrum decompositon.

Summary

- Finite value of the ration $r = N/T$ causes a huge problem in portfolio optimization.
- Cleaning is possible using a Random Matrix Model.
- Cleaning may be used to optimize a portfolio.
- The method is robust, i.e. in case the number of sectors is unknown, overestimating it does not cause any problem.
- Cleaning normally gives better results, than a simple RMM based spectrum decompositon.

Summary

- Finite value of the ration $r=N/T$ causes a huge problem in portfolio optimization.
- Cleaning is possible using a Random Matrix Model.
- Cleaning may be used to optimize a portfolio.
- The method is robust, i.e. in case the number of sectors is unknown, overestimating it does not cause any problem.
- Cleaning normally gives better results, than a simple RMM based spectrum decompositon.

Summary

- Finite value of the ration $r = N/T$ causes a huge problem in portfolio optimization.
- Cleaning is possible using a Random Matrix Model.
- Cleaning may be used to optimize a portfolio.
- The method is robust, i.e. in case the number of sectors is unknown, overestimating it does not cause any problem.
- Cleaning normally gives better results, than a simple RMM based spectrum decompositon.

Summary

- Finite value of the ration $r = N/T$ causes a huge problem in portfolio optimization.
- Cleaning is possible using a Random Matrix Model.
- Cleaning may be used to optimize a portfolio.
- The method is robust, i.e. in case the number of sectors is unknown, overestimating it does not cause any problem.
- Cleaning normally gives better results, than a simple RMM based spectrum decompositon.