Portfolio Optimization with Statistical Physics

Gábor Papp[1],
Imre Kondor[2], Fabio Caccioli[3]

May 17th, 2017, ELTE

[1]Eötvös University, Budapest,[2]Parmenides Foundation, Pullach, [3]University College, London

Portfolio Optimization with Statistical Physics

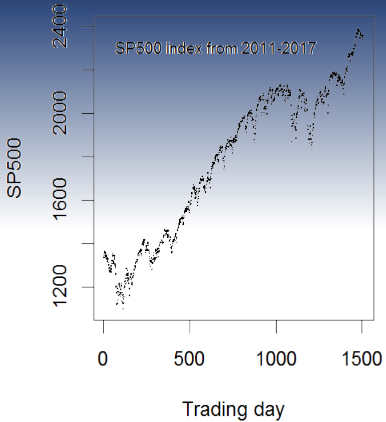
Gábor Papp[1],
Imre Kondor[2], **Fabio Caccioli**[3]

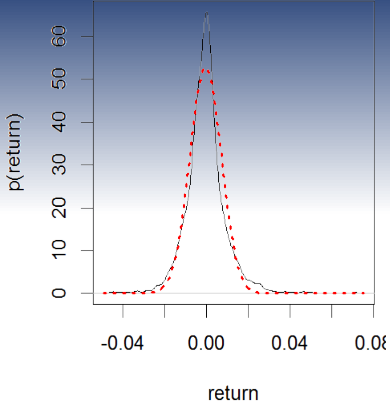
May 17th, 2017, ELTE

[1]Eötvös University, Budapest,[2]Parmenides Foundation, Pullach, [3]University College, London

Outline

- Introduction to the portfolio problem
- Markowitz solution (no constrains)
- Analytical solution: the replica trick
- Regularization / constrains
    - No short selling
- Caveats in practical applications
- Results



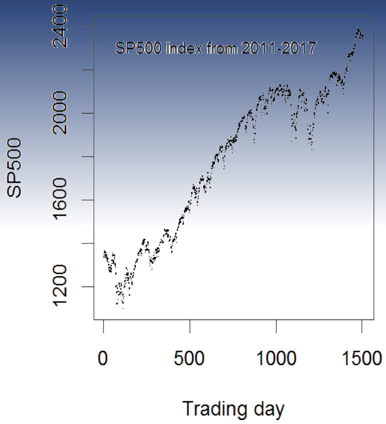


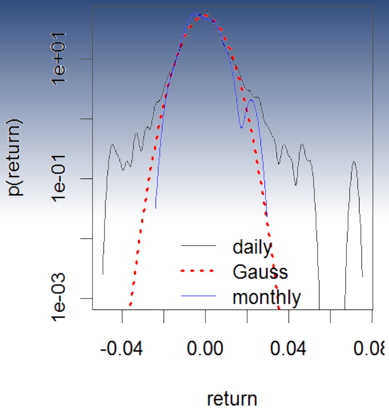
SP500 index

SP500 index return distribution

The return is defined through the price X_{it} as

$$x_{it} = \frac{X_{i,t+1} - X_{i,t}}{X_{it}} \approx \log \frac{X_{i,t+1}}{X_{i,t}}, \text{ additive}.$$





SP500 index

SP500 index return distribution

The return is defined through the price X_{it} as

$$x_{it} = \frac{X_{i,t+1} - X_{i,t}}{X_{it}} \approx \log \frac{X_{i,t+1}}{X_{i,t}}, \text{additive}.$$

The Markowitz solution

There are N assets to invest in, the covariance matrix between returns is given by $C_{ij}, i, j = 1, N$.
Find weights w_i, normalized as $\sum_i w_i = 1$, that the risk $\frac{1}{2}\sum_{ij} w_i C_{ij} w_j$ is minimal. The minimal risk solution is given by

$$w_i = \frac{\sum_j C_{ij}^{-1}}{\sum_{ij} C_{ij}^{-1}} \longrightarrow \frac{\frac{1}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}.$$

The solution involves the *inverse* of the covariance matrix, and hence, has problems, when C_{ij} is not invertible:

- two assets behave similarly (not distinguishable)
- lack of information

The Covariance Matrix

In the following we distinguish the theoretical / true covariance matrix $C_{ij}^{(0)}$ and the empirical / noisy one, C_{ij}:

$$C_{ij} = \frac{1}{T}\sum_{t=1}^{T} x_{it}x_{jt} = \frac{1}{T}XX^{\dagger},$$

where

$$X = \underbrace{\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1T} \\ x_{21} & x_{22} & \dots & x_{2T} \\ \dots & & & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NT} \end{pmatrix}}_{T \text{measurement points}} \Bigg\} N \text{ channels}$$

The Covariance Matrix

In the following we distinguish the theoretical / true covariance matrix $C_{ij}^{(0)}$ and the empirical / noisy one, C_{ij}:

$$C_{ij} = \frac{1}{T} \sum_{t=1}^{T} x_{it} x_{jt}$$

- For $T < N$ the empirical covariance matrix picks up zero modes, the optimization problem is not solvable!
- Also, as $T > N$ approaches N, the empirical covariance matrix picks up larger and larger amount of "noise", and the estimate is less and less reliable;
- As $r = N/T \to 0$ the empirical covariance matrix approaches the theoretical one.

Regularization

The error of optimization

$$q_0 = \frac{\sum_{ij} w_i C_{ij}^{(0)} w_j}{\sum_{ij} w_i^{(0)} C_{ij}^{(0)} w_j^{(0)}} \sim \frac{1}{1-r}$$

To reduce error and reach the $T < N$ region a regularization is introduced (extra knowledge about the system):

- l_2: $\sum w_i^2$ is minimized (distribute equally the weights)
- l_1: $\sum |w_i|$ is minimized (cancel the "irrelevant" weights)
- l_α: $\sum |w_i|^\alpha$ is minimized (e.g. $\alpha = 3/2$ for liquidity problems)

Regularization also appears in Deep Learning!

Optimization as statistical physical problem

The optimization problem:

$$\min_{\vec{w}}\left\{\frac{1}{2}\sum_{ij} w_i x_{it} x_{jt} w_j + g(\vec{w})\right\}, \qquad \text{s.t.} \sum_i w_i = N,$$

with the asymmetric ℓ_1 regularizer

$$g(\vec{w}) = \eta_1 \sum_i w_i \theta(w_i) - \eta_2 \sum_i w_i \theta(-w_i). \tag{1}$$

The partition function with inverse temperature $\gamma \to \infty$ is

$$Z(\vec{w}) = \left\langle \int_{-\infty}^{\infty} \prod_{i=1}^{N} \mathrm{d}w_i e^{-\gamma\left(\frac{1}{2}\sum_{i,j,t} w_i x_{it} x_{jt} w_j + g(\vec{w})\right)} \prod_a \delta\left(\sum_i w_i - N\right)\right\rangle_{\vec{x}_t}.$$

The replica trick takes care of averaging the logarithm of the partition function, based on the identity

$$\langle \log Z \rangle = \lim_{n\to 0} \frac{\partial \langle Z^n \rangle}{\partial n}.$$

$$Z_n(\vec{w}) = \left\langle \int_{-\infty}^{\infty} \prod_{i=1}^{N} \prod_{a=1}^{n} dw_i^a e^{-\gamma \left(\frac{1}{2} \sum_{i,j,t,a} w_i^a x_{it} x_{jt} w_j^a + g(\vec{w}) \right)} \prod_a \delta(\sum_i w_i^a - N) \right\rangle_{\vec{x}_t}$$

is the partition function of n replicas and equivalent to Z^n.

First, using the Hubbard-Stratonovich transformation introducing an auxiliary field ϕ_{at}, we *linearize* the exponent in x_{it}:

$$e^{-\frac{\gamma}{2} \sum_{i,j,t,a} w_i^a x_{it} x_{jt} w_j^a} = \int_{-\infty}^{\infty} \prod_{a,t} \mathrm{d}\phi_{at} \; e^{-\frac{1}{2} \sum_{a,t} \phi_{a,t}^2 + i\sqrt{\gamma} \sum_{i,t,a} \phi_{at} w_i^a x_{it}}$$

Now, the averaging over x_{it} can be done with the probability density $e^{-\frac{x_{it}^2}{2\sigma_i^2}}$.

$$-\frac{x_{it}^2}{2\sigma_i^2} + i\sqrt{\gamma}\sum_{i,t,a}\phi_{at}w_i^a x_{it} = -\frac{1}{2\sigma_i^2}\left(x_{it} - i\sqrt{\gamma}\sum_{i,t,a}\phi_{at}\sigma_i^2 w_i^a\right)^2 - \frac{\gamma}{2}\sum_{a,b,t}\phi_{at}\phi_{bt}\sum_i \sigma_i^2 w_i^a w_i^b$$

$$Q_{ab} = \frac{1}{N}\sum_i \sigma_i^2 w_i^a w_i^b \qquad \text{overlap matrix.}$$

$$Z_n(\vec{w}) = \int_{-\infty}^{\infty}\prod_{i,a,b,t}\mathrm{d}w_i^a\mathrm{d}Q_{ab}\mathrm{d}\hat{Q}_{ab}\mathrm{d}\phi_{at}\mathrm{d}\lambda^a\ e^{-\frac{1}{2}\sum_{a,t}\phi_{at}^2 - \frac{\gamma}{2}\sum_{a,b,t}\phi_{at}Q_{ab}\phi_{bt}}$$
$$\times\ e^{\sum_{a,b}\hat{Q}_{ab}\left(NQ_{ab} - \sum_i \sigma_i^2 w_i^a w_i^b\right) + \sum_a \lambda^a\left(\sum_i w_i^a - N\right) - \gamma g(\vec{w})}$$

Integrate over Φ_{at}:

$$\int_{-\infty}^{\infty} \prod_{a,t} \mathrm{d}\phi_{at} \; e^{-\frac{1}{2}\sum_{a,b,t} \phi_{at}(\delta_{ab}+\gamma Q_{ab})\phi_{bt}} = e^{-\frac{T}{2}\log\det(\delta_{ab}+\gamma Q_{ab})}$$

Replica *symmetric* ansatz:

$$\begin{array}{lll} Q_{ab} = q_0 + \Delta, & \hat{Q}_{ab} = \hat{q}_0 + \hat{\Delta}, & a = b \\ Q_{ab} = q_0, & \hat{Q}_{ab} = \hat{q}_0 & a \neq b \end{array}$$

$$\begin{vmatrix} a+b & b & \cdots & b \\ b & a+b & \cdots & b \\ \vdots & & & \vdots \\ b & b & \cdots & a+b \end{vmatrix} = a^n \left(1 + \frac{nb}{a}\right)$$

$$\log\det\left(\delta_{ab} + \gamma Q_{ab}\right) = n \log\left(1 + \gamma\Delta\right) + n\frac{\gamma q_0}{1+\gamma\Delta} \text{ as } n \to 0.$$

$$\mathrm{Tr}\hat{Q}_{ab}Q_{ab} = n\left((\hat{\Delta}+\hat{q}_0)(\Delta+q_0)+(n-1)\hat{q}_0 q_0\right)$$
$$\longrightarrow n\left(\hat{\Delta}\Delta+\hat{\Delta}q_0+\Delta\hat{q}_0\right) \text{ as } n\to 0$$

$$Z_n(\vec{w}) = \int_{-\infty}^{\infty}\prod_{i,a,b,t}\mathrm{d}w_i^a\mathrm{d}\Delta\mathrm{d}q_0\mathrm{d}\hat{\Delta}\mathrm{d}\hat{q}_0\mathrm{d}\lambda^a\ e^{-\frac{Tn}{2}\left[\log(1+\gamma\Delta)+\frac{\gamma q_0}{1+\gamma\Delta}\right]}$$
$$\times\ e^{Nn\left(\hat{\Delta}\Delta+\hat{\Delta}q_0+\Delta\hat{q}_0-\lambda\right)-\sum_{a,i}\hat{\Delta}\sigma_i^2 w_i^a w_i^a-\sum_{a,b,i}\hat{q}_0\sigma_i^2 w_i^a w_i^b+\sum_{a,i}\lambda^a w_i^a-\gamma g(\vec{w})}$$

$$-\sum_{a,i}\hat{\Delta}\sigma_i^2 w_i^a w_i^a = -Nn\hat{\Delta}\sigma^2 w^2 \qquad \sum_{a,i}\lambda^a w_i^a = nN\lambda w$$

$$-\sum_{i,a,b}\sigma_i^2\hat{q}_0 w_i^a w_i^b - \frac{1}{2}(z-\sum_{i,a}\sigma_i w_i^a\sqrt{-2\hat{q}_0})^2 = -\frac{z^2}{2}+Nnwz\sigma\sqrt{-2\hat{q}_0}$$

$$Z_n(\vec{w}) = \int_{-\infty}^{\infty} \prod_{i,a,b,t} \mathrm{d}\Delta \mathrm{d}q_0 \mathrm{d}\hat{\Delta} \mathrm{d}\hat{q}_0 \mathrm{d}\lambda \, e^{Nn\left[\frac{-1}{2r}\left(\log(1+\gamma\Delta)+\frac{\gamma q_0}{1+\gamma\Delta}\right)+\left(\hat{\Delta}\Delta+\hat{\Delta}q_0+\Delta\hat{q}_0-\lambda\right)\right]}$$

$$\times \int_{-\infty}^{\infty} \mathrm{d}w \mathrm{d}z \; e^{\int \mathrm{d}\sigma p(\sigma) Nn\left[-\hat{\Delta}\sigma^2 w^2 + wz\sigma\sqrt{-2\hat{q}_0} + \lambda w - \gamma g(\vec{w})\right] - \frac{z^2}{2}}$$

with $r = N/T$.

As for $n \to 0$, $X^n \approx 1 + n \log X \;\; \longrightarrow \log\langle X^n\rangle \simeq n\langle \log X\rangle$,

$$Z_n(\vec{w}) = \int_{-\infty}^{\infty} \prod_{i,a,b,t} \mathrm{d}\Delta \mathrm{d}q_0 \mathrm{d}\hat{\Delta} \mathrm{d}\hat{q}_0 \mathrm{d}\lambda \, e^{Nn\left[\frac{-1}{2r}\left(\log(1+\gamma\Delta)+\frac{\gamma q_0}{1+\gamma\Delta}\right)+\left(\hat{\Delta}\Delta+\hat{\Delta}q_0+\Delta\hat{q}_0-\lambda\right)\right]}$$

$$\times \, e^{Nn\left\langle \log \int_{-\infty}^{\infty} \mathrm{d}w \; \left[-\hat{\Delta}\sigma^2 w^2 + wz\sigma\sqrt{-2\hat{q}_0} + \lambda w - \gamma g(\vec{w})\right] \right\rangle_{z,\sigma}}$$

$$= \int \mathrm{d}\lambda \mathrm{d}q_0 \mathrm{d}\Delta \mathrm{d}\hat{q}_0 \mathrm{d}\hat{\Delta} \; e^{-\gamma Nnf(\lambda, q_0, \Delta, \hat{q}_0, \hat{\Delta})}$$

Free energy

$$f(\lambda, q_0, \Delta, \hat{q}_0, \hat{\Delta}) = \frac{1}{2\gamma r}\left[\log(1+\gamma\Delta) + \frac{\gamma q_0}{1+\gamma\Delta}\right] + \frac{\lambda}{\gamma}$$
$$-\frac{1}{\gamma}(\hat{q}_0\Delta + q_0\hat{\Delta} + \Delta\hat{\Delta}) - \frac{1}{\gamma}\left\langle \log\int dw\, e^{-\hat{\Delta}\sigma^2 w^2 + wz\sigma\sqrt{-2\hat{q}_0} + \lambda w - g(\vec{w})} \right\rangle_{z\sigma}$$

Performing the change of variables $\Delta \to \Delta/\gamma$, $\hat{q}_0 \to \gamma^2 \hat{q}_0$, $\hat{\Delta} \to \gamma\hat{\Delta}$, $\lambda \to \gamma\lambda$ and taking the limit $\gamma \to \infty$ we finally have in the saddle point approximation:

$$f(\lambda, q_0, \Delta, \hat{q}_0, \hat{\Delta}) = \frac{q_0}{2r(1+\Delta)} - \hat{q}_0\Delta - \hat{\Delta} q_0 + \lambda + \min_{\vec{w}}\left\langle V(\vec{w})\right\rangle_{z\sigma},$$

where

$$V = \hat{\Delta}\sigma^2 w^2 - wz\sigma\sqrt{-2\hat{q}_0} - \lambda w + \eta_1\theta(w) - \eta_2\theta(-w).$$

The minimum of the potential is at

$$w^* = \frac{\sigma z\sqrt{-2\hat{q}_0} + \lambda - \eta_1\theta(w^*) + \eta_2\theta(-w^*)}{2\hat{\Delta}\sigma^2}.$$

Substituting this back into the potential and performing the double average over z and σ

$$\langle V^*\rangle_{z\sigma} = \frac{\hat{q}_0}{\hat{\Delta}}\frac{1}{N}\sum_i \left[W\left(\frac{\lambda-\eta_1}{\sigma_i\sqrt{-2\hat{q}_0}}\right) + W\left(-\frac{\lambda+\eta_2}{\sigma_i\sqrt{-2\hat{q}_0}}\right)\right].$$

where

$$W(x)=\int_{-\infty}^{x} dt\,\Psi(t), \quad \Psi(x)=\int_{-\infty}^{x} dt\,\Phi(t), \quad \Phi(x)=\int_{-\infty}^{x} dt\,\phi(t), \quad \phi(t)=\frac{e^{-t^2/2}}{\sqrt{2\pi}}$$

Thus the free energy

$$f = \lambda - \Delta\hat{q}_0 - \hat{\Delta}q_0 + \frac{q_0}{2r(1+\Delta)} + \langle V^*\rangle_{z\sigma}$$

$$\frac{\partial f}{\partial \lambda} = \frac{\partial f}{\partial \hat{q}_0} = \frac{\partial f}{\partial \hat{\Delta}} = \frac{\partial f}{\partial q_0} = \frac{\partial f}{\partial \Delta} = 0 \quad \text{(saddle point eq.)}$$

Unconstrained case: $\eta_1 = \eta_2 = 0$

$$f = \lambda - \Delta \hat{q}_0 - \hat{\Delta} q_0 + \frac{1}{2r}\frac{q_0}{1+\Delta} + \frac{\hat{q}_0}{2\hat{\Delta}} - \frac{\lambda^2}{4\hat{\Delta}}\frac{1}{N}\sum_i \frac{1}{\sigma_i^2}.$$

with the solution:

$$\lambda = \frac{1}{\frac{1}{N}\sum_i \frac{1}{\sigma_i^2}}\frac{1-r}{r} = 2f, \qquad \text{chemical potential}$$

$$\Delta = \frac{r}{1-r}, \qquad \text{susceptibility}$$

$$q_0 = \frac{1}{\frac{1}{N}\sum_i \frac{1}{\sigma_i^2}}\frac{1}{1-r}, \qquad \text{estimation error}$$

$$\hat{q}_0 = -\frac{1}{\frac{1}{N}\sum_i \frac{1}{\sigma_i^2}}\frac{1-r}{2r},$$

$$\hat{\Delta} = \frac{1-r}{2r},$$

Weight distribution

Returning to the saddle point solution of w,

$$w* = \frac{\sigma z\sqrt{-2\hat{q}_0} + \lambda}{2\hat{\Delta}\sigma^2} \quad \text{with } \eta_1 = \eta_2 = 0$$

w can be written as

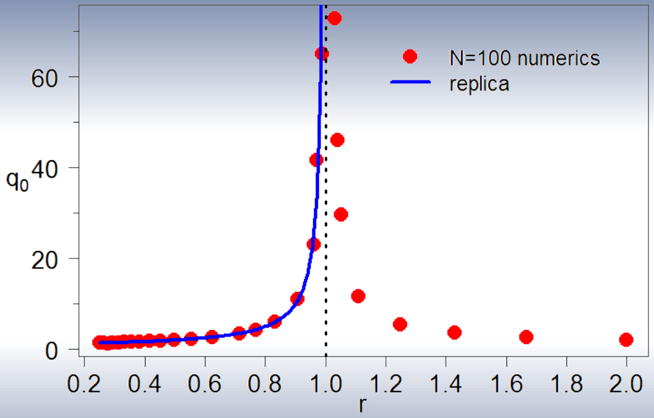
$$w_i^* = \frac{\sigma_w}{\sigma_i}\left(z - \frac{z_*}{\sigma_i}\right), \quad \sigma_w = \frac{\sqrt{-2\hat{q}_0}}{2\hat{\Delta}}, \; z_* = -\frac{\lambda}{\sqrt{-2\hat{q}_0}}$$

and since z is distributed from the normal distribution,

$$p(w) = \frac{1}{N}\sum_i \mathcal{N}\left(-\frac{\sigma_w z_*}{\sigma_i^2}, \frac{\sigma_w}{\sigma_i}\right).$$

As $r \to 0$, $p(w) \to \frac{1}{N}\sum_i \mathcal{N}(\frac{\frac{1}{\sigma_i^2}}{\frac{1}{N}\sum_i \frac{1}{\sigma_i^2}}, 0)$: Markowitz solution

Unconstrained case



Critical point, $\boxed{r_c = 1}$ is described as $\Delta \to \infty$, or $\lambda \to 0$.

No short selling: $\eta_1 = 0, \eta_2 \to \infty$

$$f = \lambda - \Delta\hat{q}_0 - \hat{\Delta} q_0 + \frac{1}{2r}\frac{q_0}{1+\Delta} + \frac{\hat{q}_0}{\hat{\Delta}}\frac{1}{N}\sum_i W\left(\frac{\lambda}{\sigma_i\sqrt{-2\hat{q}_0}}\right).$$

with the solution:

$$\begin{aligned}
\lambda &= \frac{q_0}{r(1+\Delta)^2} = 2f, \\
\Delta &= \frac{r\frac{1}{N}\sum_i \Phi\left(\frac{\sqrt{\lambda}}{\sigma_i}\right)}{1 - r\frac{1}{N}\sum_i \Phi\left(\frac{\sqrt{\lambda}}{\sigma_i}\right)}, \\
q_0 &= \lambda r(1+\Delta)^2, \\
\hat{q}_0 &= -\frac{q_0}{2r(1+\Delta)^2}, \\
\hat{\Delta} &= \frac{1}{2r(1+\Delta)}.
\end{aligned}$$

$$\frac{1}{2r} = \frac{1}{N}\sum_i W\left(\frac{\sqrt{\lambda}}{\sigma_i}\right)$$

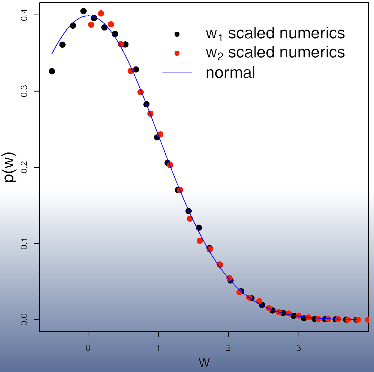
Weight distribution

$$p(w) = n_0\delta(w) + \theta(w)\frac{1}{N}\sum_i \frac{1}{\sigma_w^{(i)}\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{w-w_0^{(i)}}{\sigma_w^{(i)}}\right)^2\right]$$

$$n_0 = \frac{1}{N}\sum_i \Phi\left(-\frac{w_0^{(i)}}{\sigma_w^{(i)}}\right),$$

$$w_0^{(i)} = \frac{q_0}{(1+\Delta)}\frac{1}{\sigma_i^2},$$

$$\sigma_w^{(i)} = \frac{\sqrt{q_0 r}}{\sigma_i}.$$

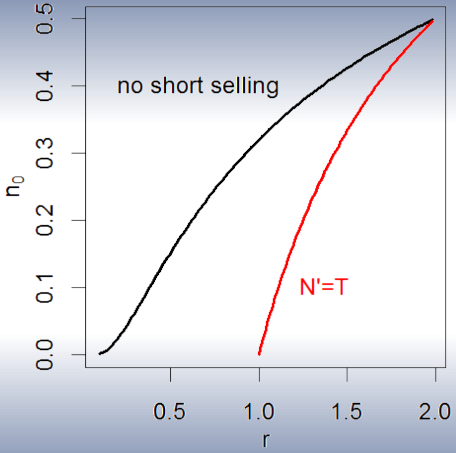


Condensation: cancelled assets

$$n_0 = \frac{1}{N} \sum_i \Phi \left(- \frac{w_0^{(i)}}{\sigma_w^{(i)}} \right),$$

is the number of cancelled assets.

Above r_c, $w_0 = 0$, that is *half* of the entries are cancelled.

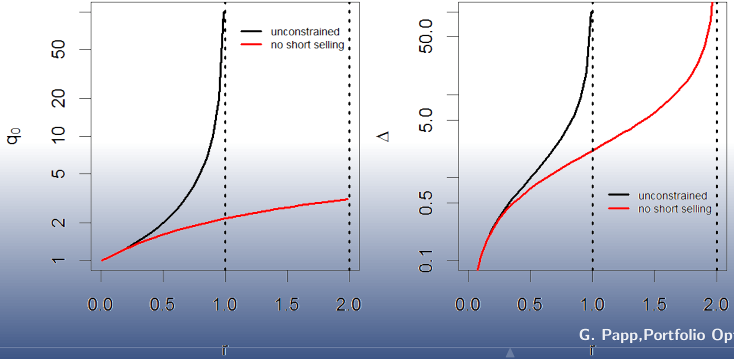


Critical point

The $\lambda = 0$ (condensation) condition for criticality with eq.

$$\frac{1}{2r} = \frac{1}{N}\sum_i W\left(\frac{\sqrt{\lambda}}{\sigma_i}\right) \to 1/4$$

yields $\boxed{r_c = 2}$. Indeed, at that point $\Phi(0) = 1/2$, and $\Delta \to \infty$.



Zero modes

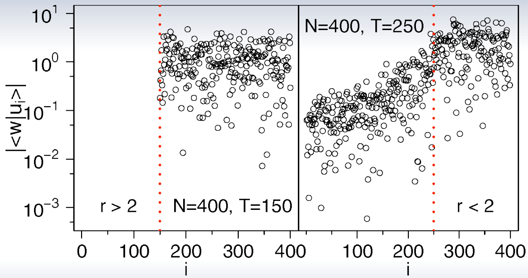
The condition to have zero risk is $\sum_{i,j,t} w_i x_{it} x_{jt} w_j = 0$, or

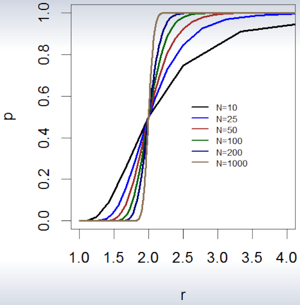
$$\sum_{t=1}^{T} \left(\sum_i w_i x_{it} \right)^2 = 0 \quad \Leftrightarrow \quad \sum_i w_i x_{it} = 0 \quad \forall t \,.$$

- For the unconstrained case as $r > 1$ zero modes of the covariance matrix automatically satisfy this condition.
- For the no short selling constrain, $w_i \geq 0$ prevents the system to spread into zero modes only, and up to $r = 2$ this condition can not be satisfied generally:
- One has to find in an N dimensional space a vector w, which is perpendicular to T random vectors, and has only positive entries:

$$p(N, T) = \frac{1}{2^{N-1}} \sum_{k=T}^{N-1} \binom{N-1}{k} \,.$$

Zero modes





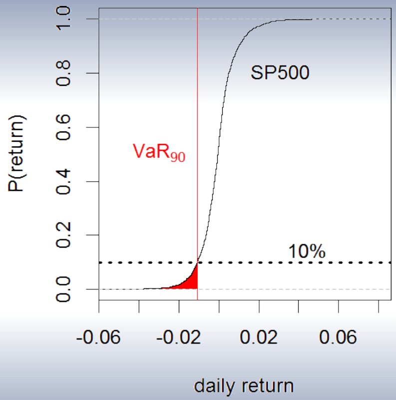
Beyond Gaussian

- Probability of riskless solution $p(N, T)$ is general, provided the underlying x distribution is symmetric and continuous;
- $r_c = 2$ is distribution independent;
- Going beyond Gaussian analytically may be possible for special distributions;
- Numerical study is possible (Student): the general behavior is similar, Gaussian seems to be the most well behaving distribution.

Conclusions

- Portolio in high-dimensional limit with fixed $r = N/T$ ratio is considered, returns from normal distribution;
- unconstrained case:
  - optimal weight distribution;
  - $r_c = 1$ recovered, connection to zero modes, distribution independence;
- constrains: no short selling
  - first analytical solution of the problem, supported by numerical calculations;
  - $r_c = 2$ found, connection to zero modes, distribution independence;
- Standard programs automatically make l_2 regularization in the zero mode sector, which leads to a dangerous solution not indicating the criticality of the system.

Other risk measures



- VaR_α: Value at Risk
  - $P(\mathrm{V}aR_\alpha) = 1 - \alpha$
  - NOT a coherent measure
- ES_α: Expected Shortfall (CVaR)
  - $\mathrm{E}S_\alpha = \frac{1}{\alpha} \int_0^\alpha \mathrm{d}\gamma \mathrm{V}aR_\gamma$
  - coherent measure

Expected Shortfall

The probability of loss $l(\{w_i\},\{x_i\}) = -X$ to be smaller than a threshold l_0 is

$$
\begin{aligned}
P(\{w_i\}, l_0) &= \int \prod_i \mathrm{d}x_i p(\{x_i\})\, \theta\left(l_0 - l(\{w_i\},\{x_i\})\right) \\
\mathrm{VaR}_\alpha &= \min\left\{l_0 : P(\{w_i\}, l_0) \geq \alpha\right\} \\
(1-\alpha)\mathrm{ES}_\alpha &= \int \prod_i \mathrm{d}x_i p(\{x_i\}) l(\{w_i\},\{x_i\}\, \theta\left(l(\{w_i\},\{x_i\}) - \mathrm{VaR}_\alpha\right)
\end{aligned}
$$

Instead, Rockafellar and Uryasev proposed

$$
\begin{aligned}
F_\alpha(\{w_i\},\epsilon) &= \epsilon + \frac{1}{1-\alpha}\int \Pi_i dx_i p(\{x_i\})\left[\ell(\{w_i\},\{x_i\}) - \epsilon\right]^+ \\
\mathrm{ES}(\{w_i\}) &= \mathrm{min}_\epsilon F_\alpha(\{w_i\},\epsilon)
\end{aligned}
$$

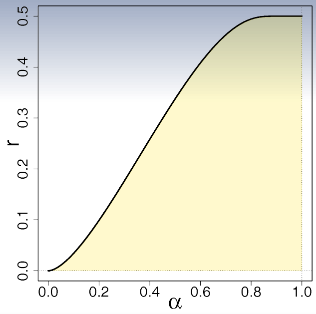
with $[x]^+ = (x + |x|)/2$.

This problem may be linearized, searching for the minimum of

$$\begin{aligned} E(\epsilon, \{u_t\}) &= (1-\alpha) T\epsilon + \sum_{t=1}^{T} u_t \\ u_t &\geq 0 \qquad \forall\ t, \\ u_t + \epsilon + \sum_{i=1}^{N} x_{it} w_i &\geq 0 \qquad \forall\ t, \end{aligned}$$

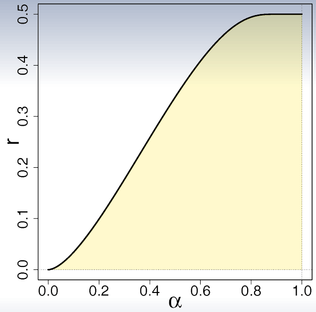
and the partition function modifies accordingly,

$$Z_\gamma\left[\{x_{i,t}\}\right] = \int_0^\infty \prod_{i=1}^{T} \mathrm{d}u_t \int_{-\infty}^{\infty} \mathrm{d}\epsilon\, \theta\left(u_t + \epsilon + \sum_{i=1}^{N} x_{i,t} w_i\right) e^{-\gamma E[\epsilon,\{u_t\}]}.$$



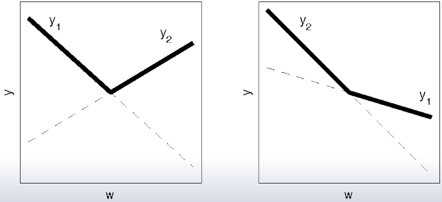
$\alpha = 1$ is the minimax risk measure.
For $N = 2$, $T = 2$:

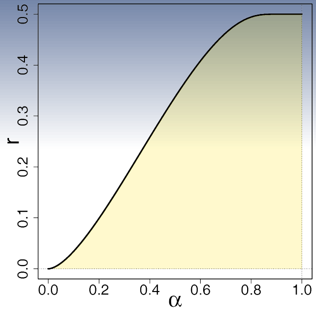
$$
\begin{aligned}
y_1 &= -wx_{11} - (1-w)x_{21} = w(x_{21} - x_{11}) \\
y_2 &= -wx_{12} - (1-w)x_{22} = w(x_{22} - x_{12})
\end{aligned}
$$



$\alpha = 1$ is the minimax risk measure.
For $N = 2$, $T = 2$:

$$y_1 = -wx_{11} - (1-w)x_{21} = w(x_{21} - x_{11})$$
$$y_2 = -wx_{12} - (1-w)x_{22} = w(x_{22} - x_{12})$$

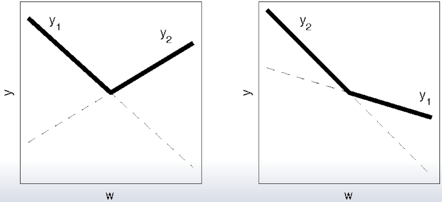




$\alpha = 1$ is the minimax risk measure.

For $N = 2$, $T = 2$:

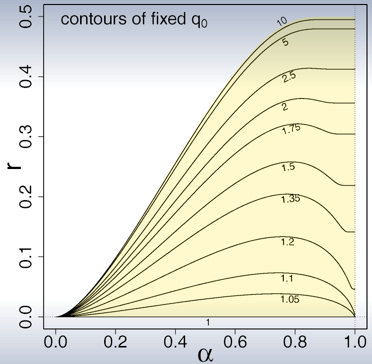
$$y_1 = -wx_{11} - (1-w)x_{21} = w(x_{21} - x_{11})$$
$$y_2 = -wx_{12} - (1-w)x_{22} = w(x_{22} - x_{12})$$



$$p = \frac{1}{2^{T-1}} \sum_{k=N-1}^{T-1} \binom{T-1}{k}.$$

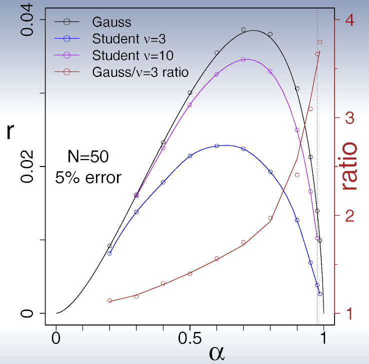
Result depends ONLY on geometry!

Error of estimation for ES



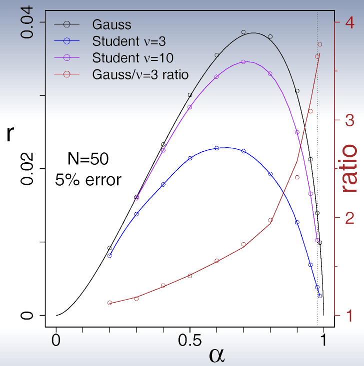
for $\alpha = 0.975$ 5% precision requires $T/N = 71$, for 20% precision $T/N = 16$, and even 50% precision is $T/N = 5$.

Error of estimation for ES



5% estimation error contour line from numerical simulations.

Error of estimation for ES



5% estimation error contour line from numerical simulations.

Introducing l_1 regularization has similar effect, than for variance+l_1: the critical line move to $2r_c^{(0)}$.