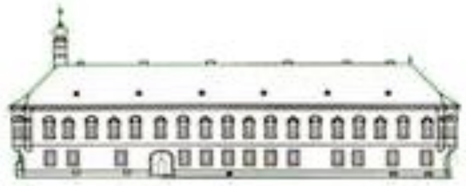


Critical Phenomena in Portfolio Selection

Imre Kondor

Collegium Budapest and Eötvös University, Budapest

Workshop on Statistical Physics and Financial Markets

April 20-21, 2007

ICTP, Trieste and EU-NEST project COMPLEXMARKETS

Contents

The subject of the talk lies at the crossroads of finance, statistical physics and computer science

The main message:

- portfolio selection is highly unstable,
- the estimation error diverges for a critical value of the ratio of the portfolio size *N* and the length of the time series *T*,
- this divergence is an algorithmic phase transition that is characterized by universal scaling laws,
- similar critical phenomena are found in other, related problems (complex optimization, multivariate regression, etc.)

Coworkers

- **Szilárd Pafka** (Paycom.net, California)
- **Gábor Nagy** (Debrecen University PhD student and CIB Bank, Budapest)
- **Nándor Gulyás** (ELTE PhD student and Collegium Budapest)
- **István Varga-Haszonits** (ELTE PhD student and Morgan-Stanley Fixed Income)
- **Andrea Ciliberti** (Roma)
- **Marc Mézard** (Orsay)
- **Stefan Thurner** (Vienna)

Rational portfolio selection seeks a tradeoff between risk and reward

- In this talk I will focus on equity portfolios

- Financial reward can be measured in terms of the return (relative gain):

$$\frac{S_t - S_{t-1}}{S_{t-1}}$$

or logarithmic return:

$$\ln\left(\frac{S_t}{S_{t-1}}\right)$$

- The characterization of risk is more controversial

The most obvious choice for a risk measure: Variance

- Its use for a risk measure assumes that the probability distribution of returns is sufficiently concentrated around the average, that there are no large fluctuations
- This is true in several instances, but we often encounter „fat tails”, huge deviations with a non-negligible probability

The most obvious choice for a risk measure: Variance

- Its use for a risk measure assumes that the probability distribution of returns is sufficiently concentrated around the average, that there are no large fluctuations
- This is true in several instances, but we often encounter „fat tails”, huge deviations with a non-negligible probability

Alternative risk measures

- There are several alternative risk measures in use in the academic literature, practice, and regulation
- Value at risk (VaR): the best among the *p%* worst losses (not convex, punishes diversification)
- Mean absolute deviation (MAD): Algorithmics
- Coherent risk measures (promoted by academics):

  Expected shortfall (ES): average loss beyond a high threshold

  Maximal loss (ML): the single worst case

Alternative risk measures

- There are several alternative risk measures in use in the academic literature, practice, and regulation
- Value at risk (VaR): the best among the *p%* worst losses (not convex, punishes diversification)
- Mean absolute deviation (MAD): Algorithmics
- Coherent risk measures (promoted by academics):

  Expected shortfall (ES): average loss beyond a high threshold

  Maximal loss (ML): the single worst case

Alternative risk measures

- There are several alternative risk measures in use in the academic literature, practice, and regulation
- Value at risk (VaR): the best among the *p%* worst losses (not convex, punishes diversification)
- Mean absolute deviation (MAD): Algorithmics
- Coherent risk measures (promoted by academics):

  Expected shortfall (ES): average loss beyond a high threshold

  Maximal loss (ML): the single worst case

Alternative risk measures

- There are several alternative risk measures in use in the academic literature, practice, and regulation
- Value at risk (VaR): the best among the *p%* worst losses (not convex, punishes diversification)
- Mean absolute deviation (MAD): Algorithmics
- Coherent risk measures (promoted by academics):

  Expected shortfall (ES): average loss beyond a high threshold

  Maximal loss (ML): the single worst case

Portfolios

- A portfolio is a linear combination (a weighted average) of assets r_i :

$$r_P = \sum_i w_i r_i$$

with a set of weights w_i that add up to unity (the budget constraint):

$$\sum_i w_i = 1$$

- The weights are not necessarily positive – short selling
- The fact that the weights can be arbitrary means that the region over which we are trying to determine the optimal portfolio is not bounded

Portfolios

- A portfolio is a linear combination (a weighted average) of assets r_i:

$$r_P = \sum_i w_i r_i$$

with a set of weights *wi* that add up to unity (the budget constraint):

$$\sum_i w_i = 1$$

- The weights are not necessarily positive – short selling
- The fact that the weights can be arbitrary means that the region over which we are trying to determine the optimal portfolio is not bounded

Portfolios

- A portfolio is a linear combination (a weighted average) of assets r_i :

$$r_P = \sum_i w_i r_i$$

with a set of weights *wi* that add up to unity (the budget constraint):

$$\sum_i w_i = 1$$

- The weights are not necessarily positive – short selling
- The fact that the weights can be arbitrary means that the region over which we are trying to determine the optimal portfolio is not bounded

Markowitz' portfolio selection theory

Rational portfolio selection realizes the tradeoff between risk and reward by minimizing the risk functional

$$\sigma_P^2 = \sum_{i,j} w_i \sigma_{ij} w_j$$

over the weights, given the expected return, the budget constraint, and possibly other costraints.

How do we know the returns and the covariances?

- In principle, from observations on the market
- If the portfolio contains N assets, we need $O(N^2)$ data
- The input data come from T observations for N assets
- The estimation error is negligible as long as $NT >> N^2$, i.e. $N << T$
- This condition is often violated in practice

How do we know the returns and the covariances?

- In principle, from observations on the market
- If the portfolio contains N assets, we need $O(N^2)$ data
- The input data come from T observations for N assets
- The estimation error is negligible as long as $NT \gg N^2$, i.e. $N \ll T$
- This condition is often violated in practice

How do we know the returns and the covariances?

- In principle, from observations on the market
- If the portfolio contains N assets, we need $O(N^2)$ data
- The input data come from T observations for N assets
- The estimation error is negligible as long as $NT>>N^2$, i.e. $N<<T$
- This condition is often violated in practice

How do we know the returns and the covariances?

- In principle, from observations on the market
- If the portfolio contains N assets, we need $O(N^2)$ data
- The input data come from T observations for N assets
- The estimation error is negligible as long as $NT \gg N^2$, i.e. $N \ll T$
- This condition is often violated in practice

How do we know the returns and the covariances?

- In principle, from observations on the market
- If the portfolio contains N assets, we need $O(N^2)$ data
- The input data come from T observations for N assets
- The estimation error is negligible as long as $NT \gg N^2$, i.e. $N \ll T$
- This condition is often violated in practice

Information deficit

- Thus the Markowitz problem suffers from the „curse of dimensions”, or from information deficit
- The estimates will contain error and the resulting portfolios will be suboptimal
- How serious is this effect?
- How sensitive are the various risk measures to this kind of error?
- How can we reduce the error?

Information deficit

- Thus the Markowitz problem suffers from the „curse of dimensions", or from information deficit
- The estimates will contain error and the resulting portfolios will be suboptimal
- How serious is this effect?
- How sensitive are the various risk measures to this kind of error?
- How can we reduce the error?

Information deficit

- Thus the Markowitz problem suffers from the „curse of dimensions”, or from information deficit
- The estimates will contain error and the resulting portfolios will be suboptimal
- How serious is this effect?
- How sensitive are the various risk measures to this kind of error?
- How can we reduce the error?

Information deficit

- Thus the Markowitz problem suffers from the „curse of dimensions”, or from information deficit
- The estimates will contain error and the resulting portfolios will be suboptimal
- How serious is this effect?
- How sensitive are the various risk measures to this kind of error?
- How can we reduce the error?

Information deficit

- Thus the Markowitz problem suffers from the „curse of dimensions", or from information deficit
- The estimates will contain error and the resulting portfolios will be suboptimal
- How serious is this effect?
- How sensitive are the various risk measures to this kind of error?
- How can we reduce the error?

Fighting the curse of dimensions

- Economists have been struggling with this problem for ages. Since the root of the problem is lack of sufficient information, the remedy is to inject external info into the estimate. This means imposing some structure on **σ**. This introduces bias, but beneficial effect of noise reduction may compensate for this.
- Examples:
  - single-factor models (β's)
  - multi-factor models
  - grouping by sectors
  - principal component analysis
  - Bayesian shrinkage estimators, etc.
  - Random matrix theory

All these help to various degrees. Most studies are based on empirical data

Our approach:

- Analytical: Applying the methods of statistical physics (random natrix theory, phase transition theory, replicas, etc.)

- Numerical: To test the noise sensitivity of various risk measures we use simulated data

  The rationale is that in order to be able to compare the sensitivity of various risk measures to noise, we better get rid of other sources of uncertainty, like non-stationarity. This can be achieved by using artificial data where we have total control over the underlying stochastic process.

  For simplicity, we mostly use iid normal variables in the following.

Our approach:

- Analytical: Applying the methods of statistical physics (random natrix theory, phase transition theory, replicas, etc.)

- Numerical: To test the noise sensitivity of various risk measures we use simulated data

  The rationale is that in order to be able to compare the sensitivity of various risk measures to noise, we better get rid of other sources of uncertainty, like non-stationarity. This can be achieved by using artificial data where we have total control over the underlying stochastic process.

  For simplicity, we mostly use iid normal variables in the following.

Our approach:

- Analytical: Applying the methods of statistical physics (random natrix theory, phase transition theory, replicas, etc.)

- Numerical: To test the noise sensitivity of various risk measures we use simulated data

  The rationale is that in order to be able to compare the sensitivity of various risk measures to noise, we better get rid of other sources of uncertainty, like non-stationarity. This can be achieved by using artificial data where we have total control over the underlying stochastic process.

  For simplicity, we mostly use iid normal variables in the following.

Our approach:

- Analytical: Applying the methods of statistical physics (random natrix theory, phase transition theory, replicas, etc.)

- Numerical: To test the noise sensitivity of various risk measures we use simulated data

  The rationale is that in order to be able to compare the sensitivity of various risk measures to noise, we better get rid of other sources of uncertainty, like non-stationarity. This can be achieved by using artificial data where we have total control over the underlying stochastic process.

  For simplicity, we mostly use iid normal variables in the following.

- For such simple underlying processes the exact risk measure can be calculated.
- To construct the empirical risk measure

$$\sigma_{ij}^{(1)} = \frac{1}{T}\sum_{t=1}^{T} y_{it} y_{jt} .$$

we generate long time series, and cut out segments of length *T* from them, as if making observations on the market.

- From these „observations" we construct the empirical risk measure and optimize our portfolio under it.

- For such simple underlying processes the exact risk measure can be calculated.
- To construct the empirical risk measure

$$\sigma_{ij}^{(1)} = \tfrac{1}{T}\sum_{t=1}^{T} y_{it} y_{jt}.$$

we generate long time series, and cut out segments of length *T* from them, as if making observations on the market.

- From these „observations" we construct the empirical risk measure and optimize our portfolio under it.

- For such simple underlying processes the exact risk measure can be calculated.
- To construct the empirical risk measure

$$\sigma_{ij}^{(1)} = \frac{1}{T}\sum_{t=1}^{T} y_{it} y_{jt} \cdot$$

we generate long time series, and cut out segments of length *T* from them, as if making observations on the market.

- From these „observations" we construct the empirical risk measure and optimize our portfolio under it.

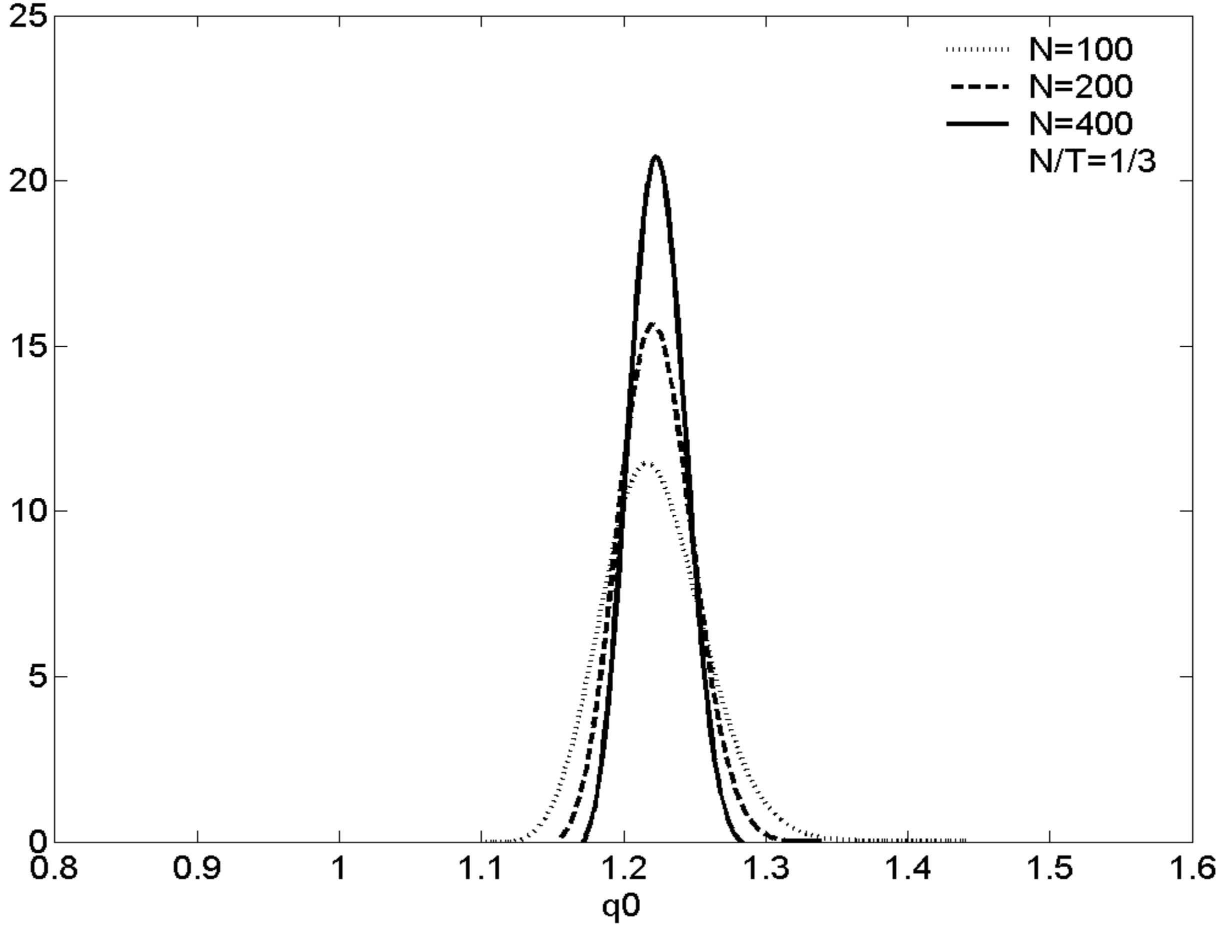
The ratio q_o of the empirical and the exact risk measure is a measure of the estimation error due to noise:

$$q_0^2 = \frac{\sum_{ij} w_i^{(1)*} \sigma_{ij}^{(0)} w_j^{(1)*}}{\sum_{ij} w_i^{(0)*} \sigma_{ij}^{(0)} w_j^{(0)*}}$$

The case of variance as a risk measure

- The relative error q_0 of the optimal portfolio is a random variable, fluctuating from sample to sample.
- The weights of the optimal portfolio also fluctuate.

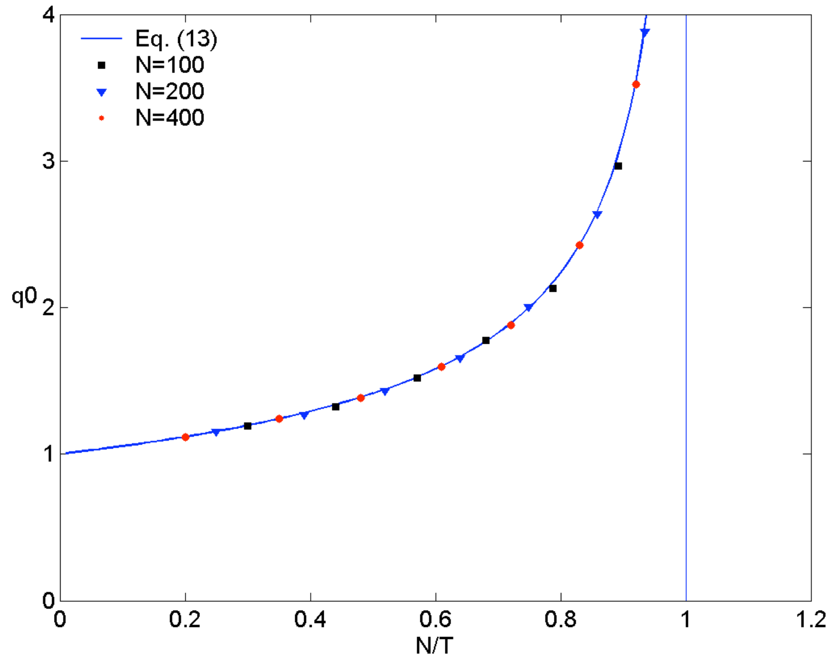
The distribution of *qo* over the samples



Critical behaviour for *N*,*T* large, with *N/T*=fixed

The average of q_o as a function of *N/T* can be calculated from random matrix theory: it diverges at the critical point *N/T*=1

$$q_0 = \frac{1}{\sqrt{1-\frac{N}{T}}}$$



Associated statistical physics model: a random Gaussian model

$$\mathrm{H} = \sum_{i,j=1}^{N} \sigma_{i,j} \cdot w_i \cdot w_j + \lambda \cdot \sum_{i=1}^{N} w_i + \mu \cdot \sum_{i=1}^{N} w_i^2$$

$$\sigma_{i,j} = \frac{1}{T} \cdot \sum_{t=1}^{T} y_{i,t} \cdot y_{j,t}$$

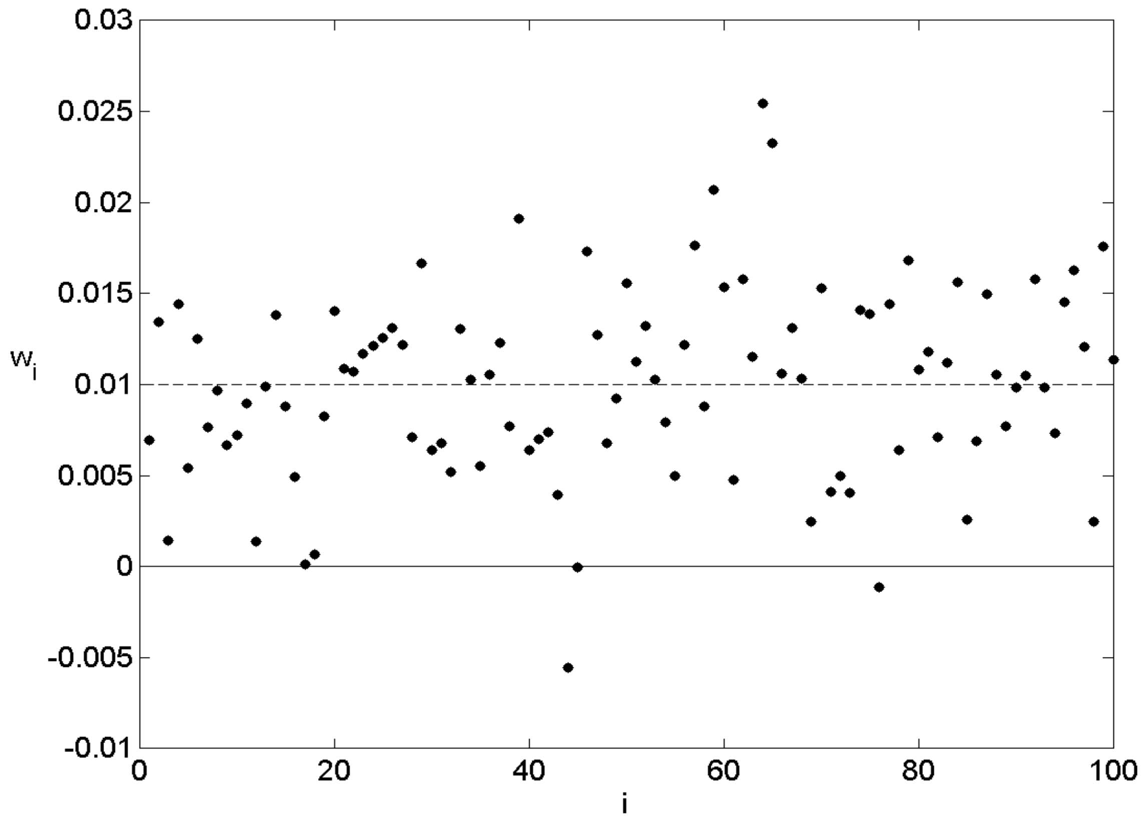
The standard deviation of the estimation error diverges even more strongly than the average:

$$\sigma(q_0) = \frac{const}{\sqrt{N} \cdot (1-r)} \quad , \quad \text{where } r = N/T$$

$$const \approx 0.67$$

Instability of the weigths

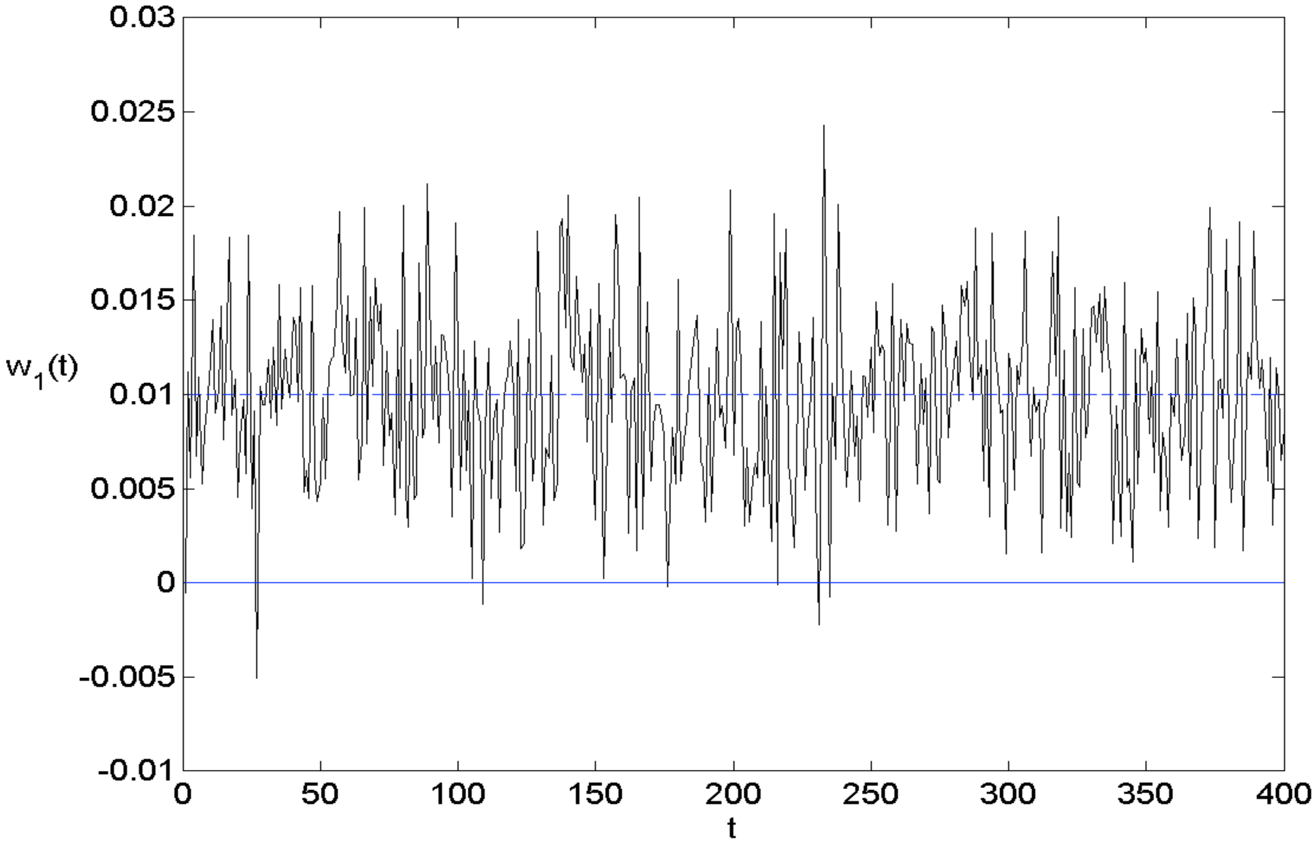
The weights of a portfolio of N=100 iid normal variables for a given sample, T=500



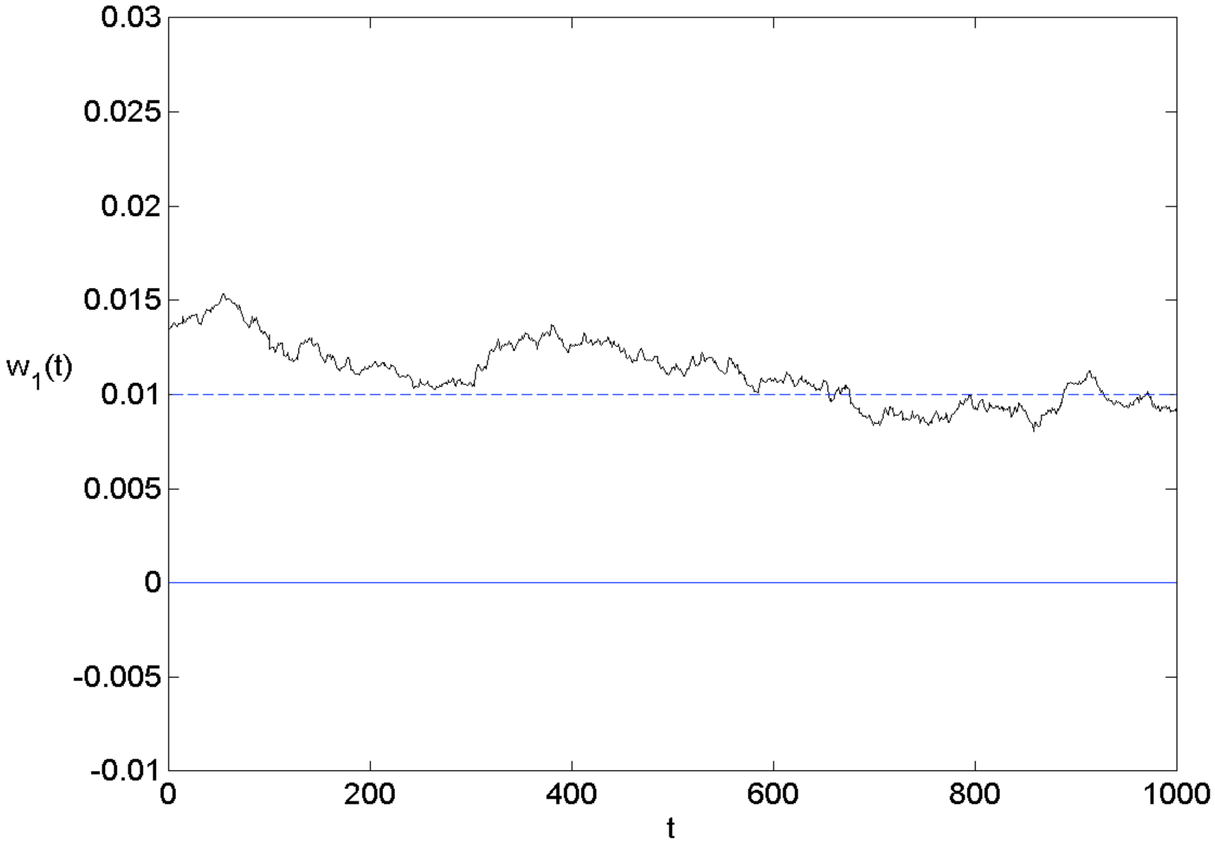
The distribution of weights in a given sample

- The optimization hardly determines the weights even far from the critical point!
- The standard deviation of the weights relative to their exact average value also diverges at the critical point

Fluctuations of a given weight from sample to sample, non-overlapping time-windows, *N*=100, *T*=500



Fluctuations of a given weight from sample to sample, time-windows shifted by one step at a time, *N*=100, *T*=500



If short selling is banned

If the weights are constrained to be positive, the instability will manifest itself by more and more weights becoming zero – the portfolio spontaneously reduces its size!

Explanation: the solution would like to run away, the constraints prevent it from doing so, therefore it will stick to the walls.

Similar effects are observed if we impose any other linear constraints, like limits on sectors, etc.

It is clear, that in these cases the solution is determined more by the constraints than the objective function.

If short selling is banned

If the weights are constrained to be positive, the instability will manifest itself by more and more weights becoming zero – the portfolio spontaneously reduces its size!

Explanation: the solution would like to run away, the constraints prevent it from doing so, therefore it will stick to the walls.

Similar effects are observed if we impose any other linear constraints, like limits on sectors, etc.

It is clear, that in these cases the solution is determined more by the constraints than the objective function.

If short selling is banned

If the weights are constrained to be positive, the instability will manifest itself by more and more weights becoming zero – the portfolio spontaneously reduces its size!

Explanation: the solution would like to run away, the constraints prevent it from doing so, therefore it will stick to the walls.

Similar effects are observed if we impose any other linear constraints, like limits on sectors, etc.

It is clear, that in these cases the solution is determined more by the constraints than the objective function.

If short selling is banned

If the weights are constrained to be positive, the instability will manifest itself by more and more weights becoming zero – the portfolio spontaneously reduces its size!

Explanation: the solution would like to run away, the constraints prevent it from doing so, therefore it will stick to the walls.

Similar effects are observed if we impose any other linear constraints, like limits on sectors, etc.

It is clear, that in these cases the solution is determined more by the constraints than the objective function.

If the variables are not iid

Experimenting with various market models (one-factor, market plus sectors, positive and negative covariances, etc.) shows that the main conclusion does not change – a manifestation of universality

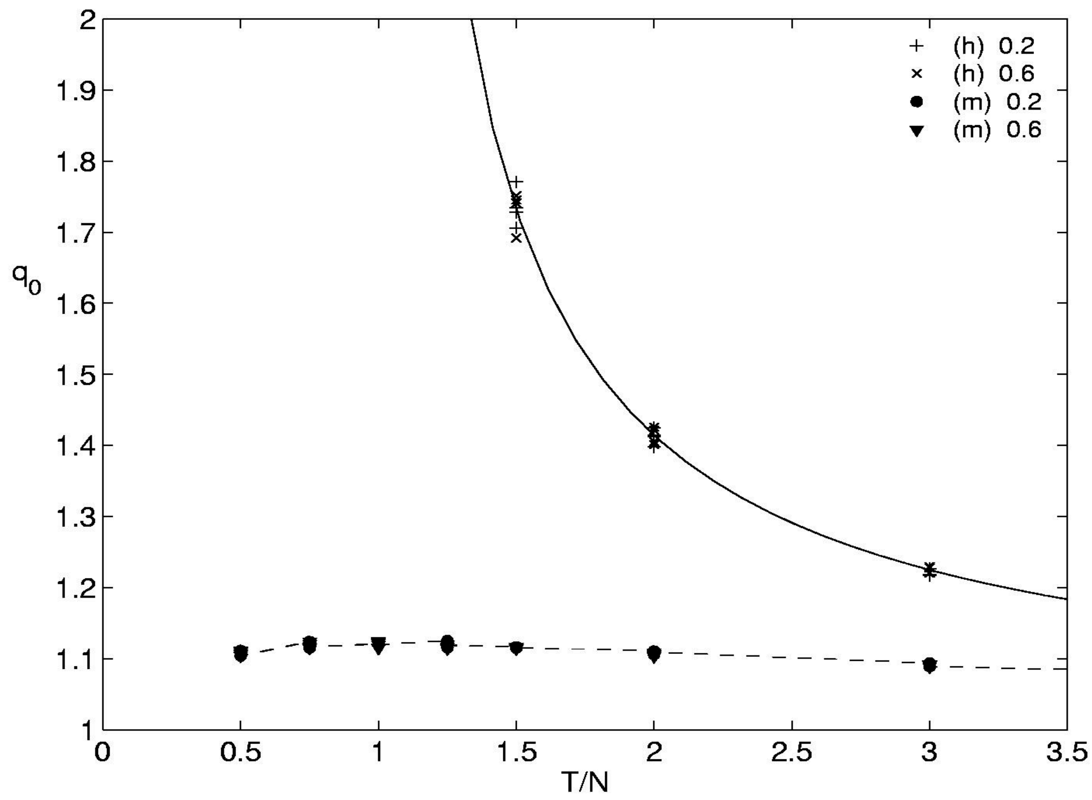
Overwhelmingly positive correlations tend to enhance the instability, negative ones decrease it, but they do not change the power of the divergence, only its prefactor

If the variables are not iid

Experimenting with various market models (one-factor, market plus sectors, positive and negative covariances, etc.) shows that the main conclusion does not change – a manifestation of universality.

Overwhelmingly positive correlations tend to enhance the instability, negative ones decrease it, but they do not change the power of the divergence, only its prefactor

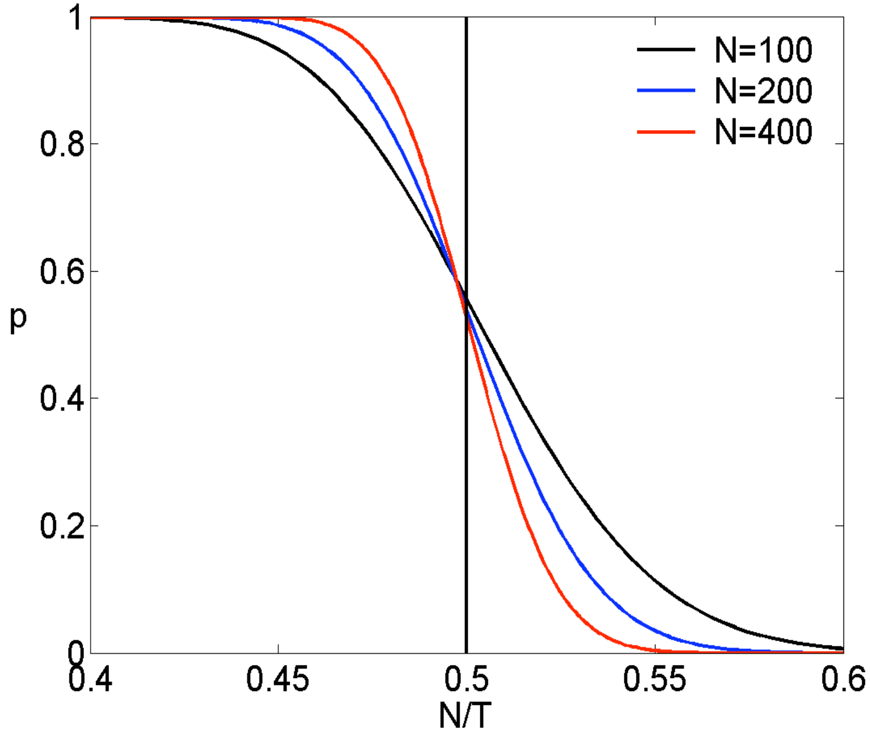
After filtering the noise is much reduced, and we can even penetrate into the region below the critical point $T<N$. BUT: the weights remain extremely unstable even after filtering



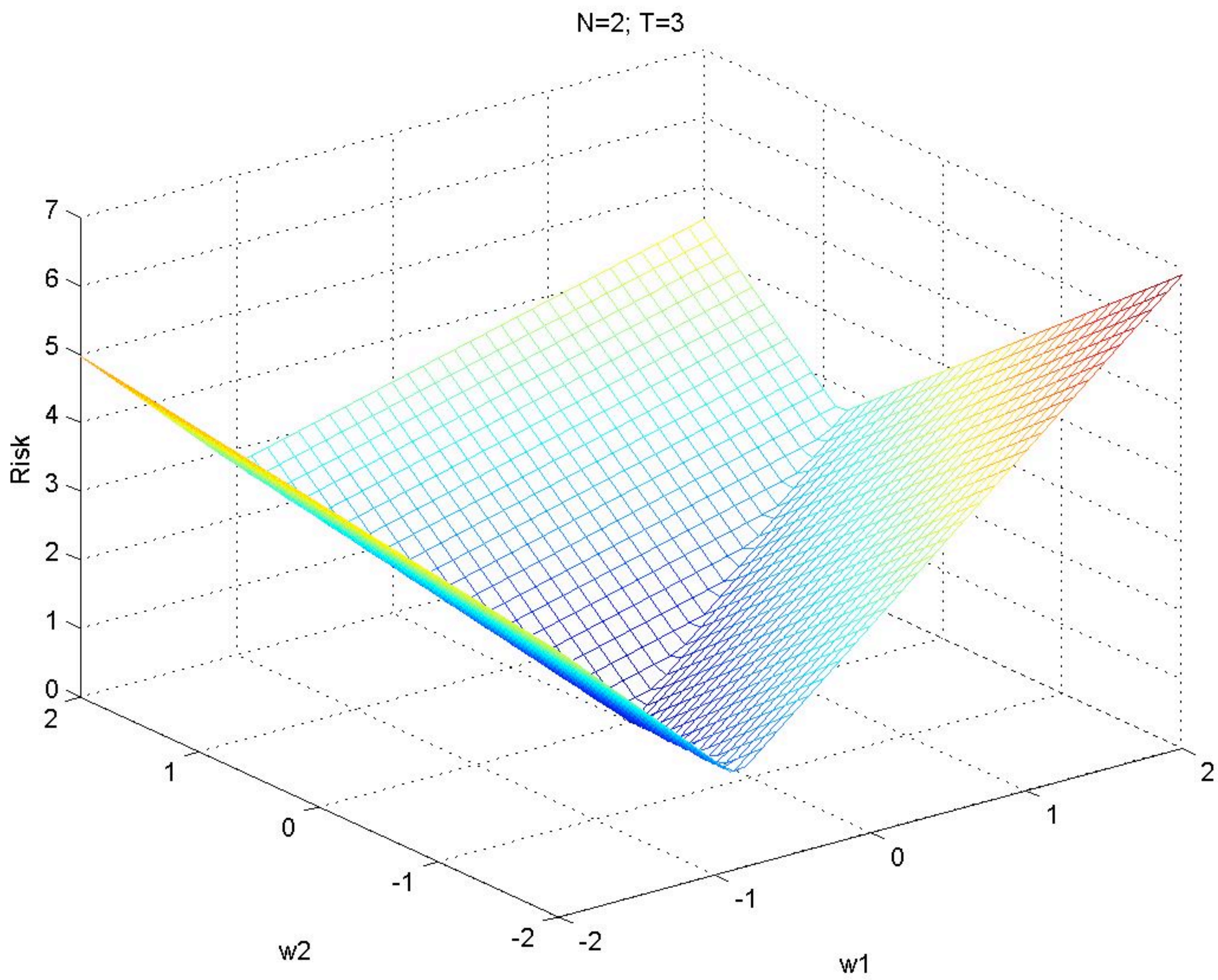
Similar studies under mean absolute deviation, expected shortfall and maximal loss

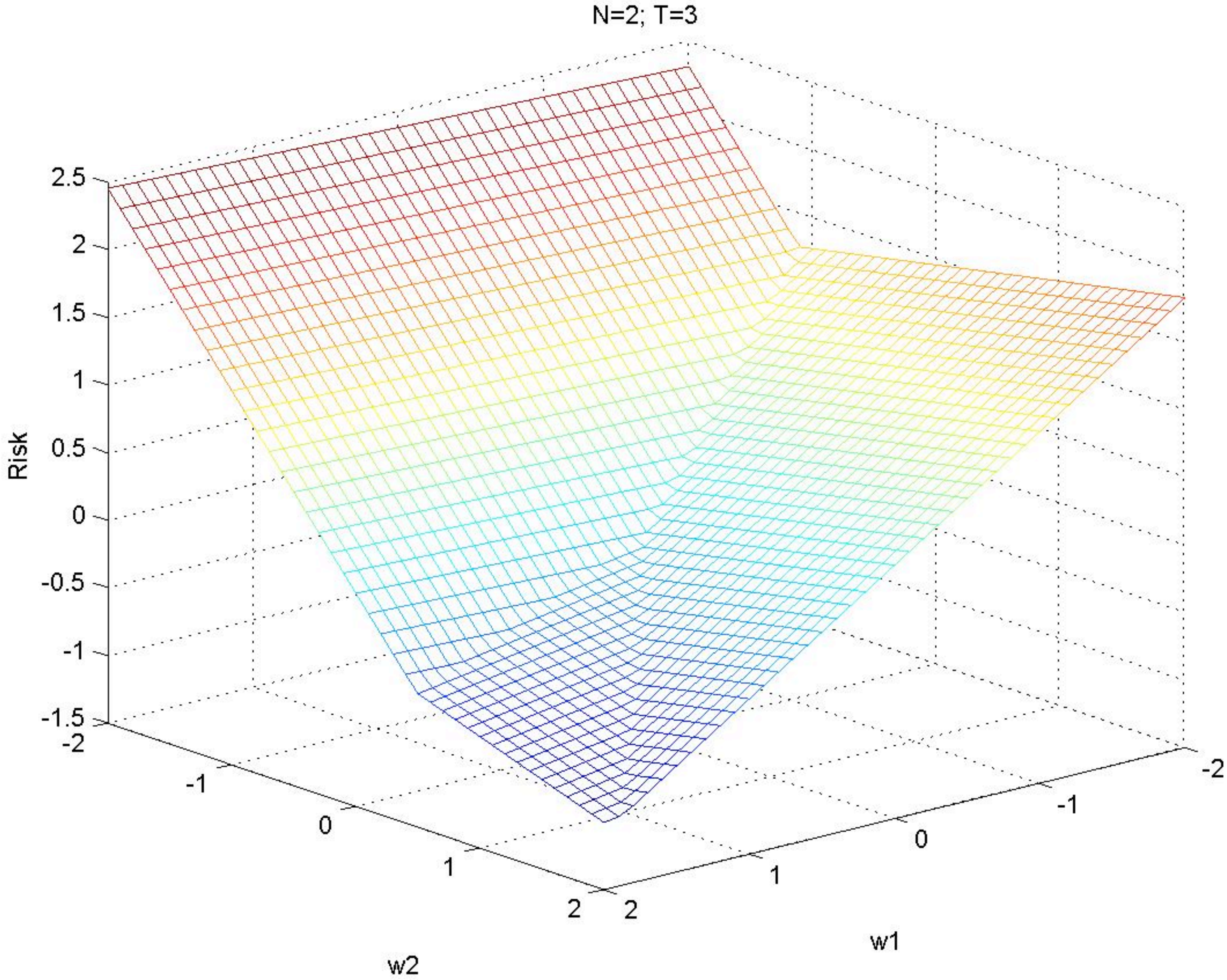
- Lead to similar conclusions, except that the effect of estimation error is even more serious
- In addition, no convincing filtering methods exist for these measures
- In the case of coherent measures the existence of a solution becomes a probabilistic issue, depending on the sample
- Calculation of this probability leads to some intriguing problems in random geometry

Probability of finding a solution for the minimax problem:

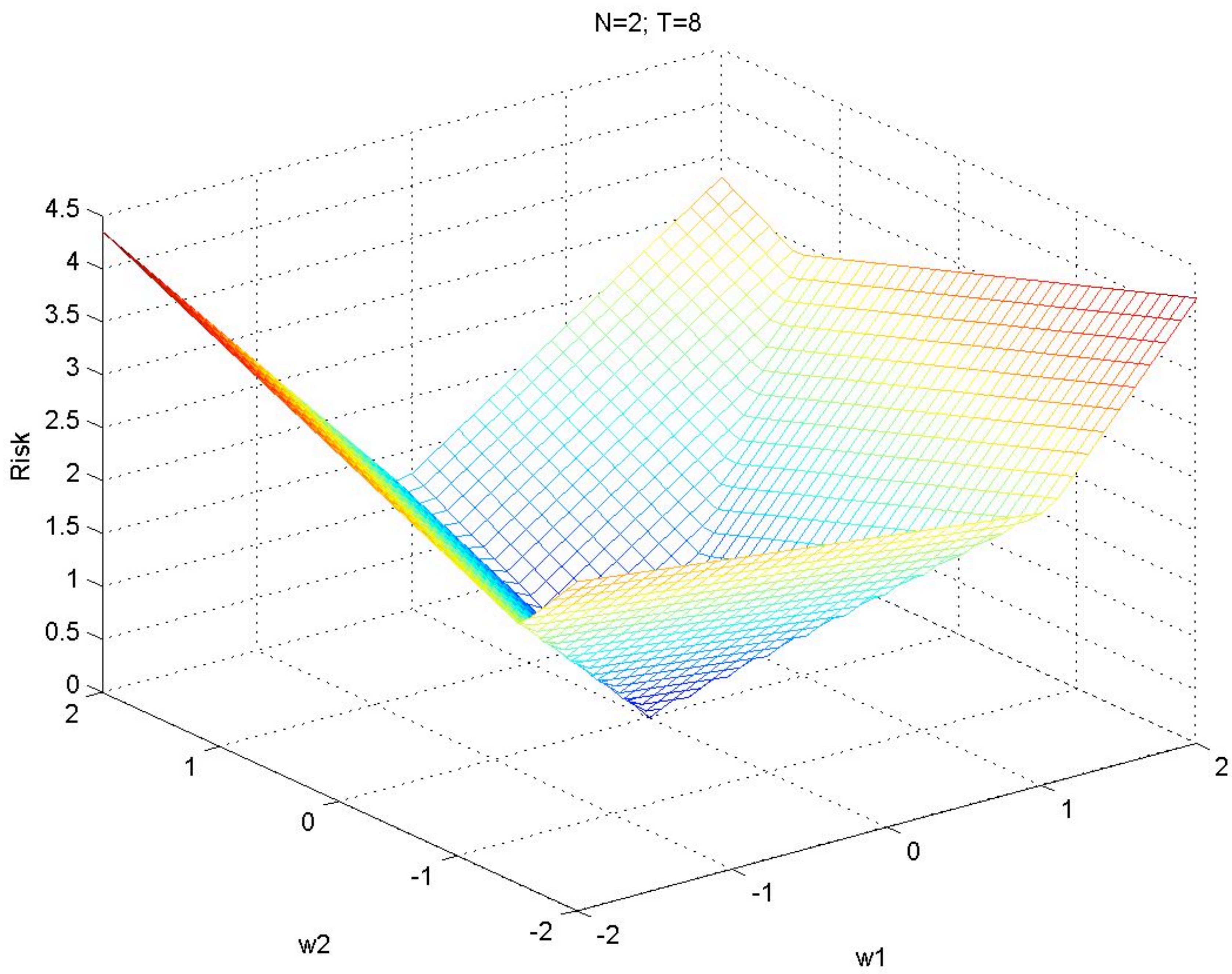


$$p = \frac{1}{2^{T-1}} \sum_{k=N-1}^{T-1} \binom{T-1}{k}$$

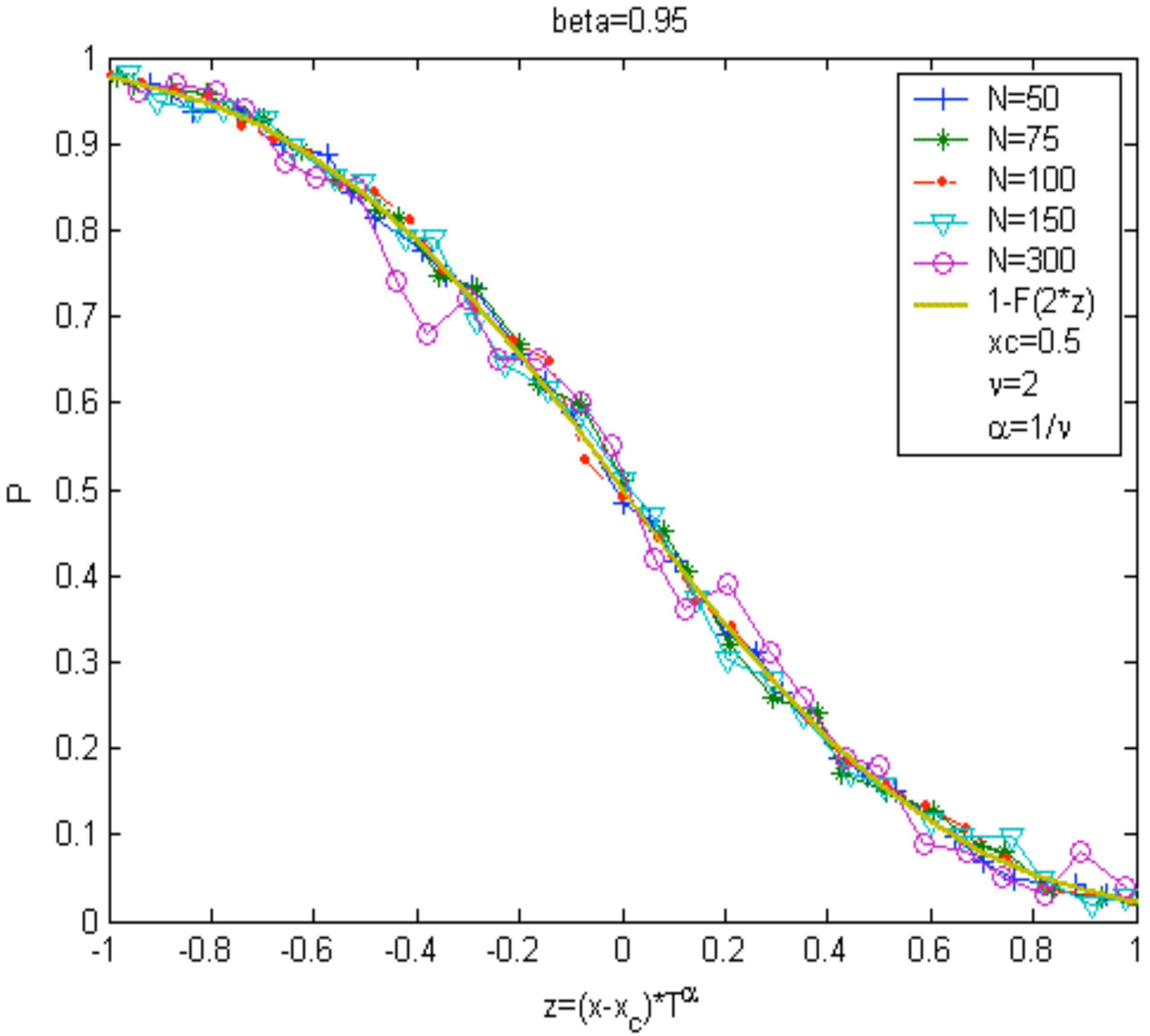

N=2; T=3
Risk
0
1
2
3
4
5
6
7
w2
2
1
0
-1
-2
w1
-2
-1
0
1
2



N=2; T=3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Risk
-2
-1
0
1
2
w2
-2
-1
0
1
2
w1

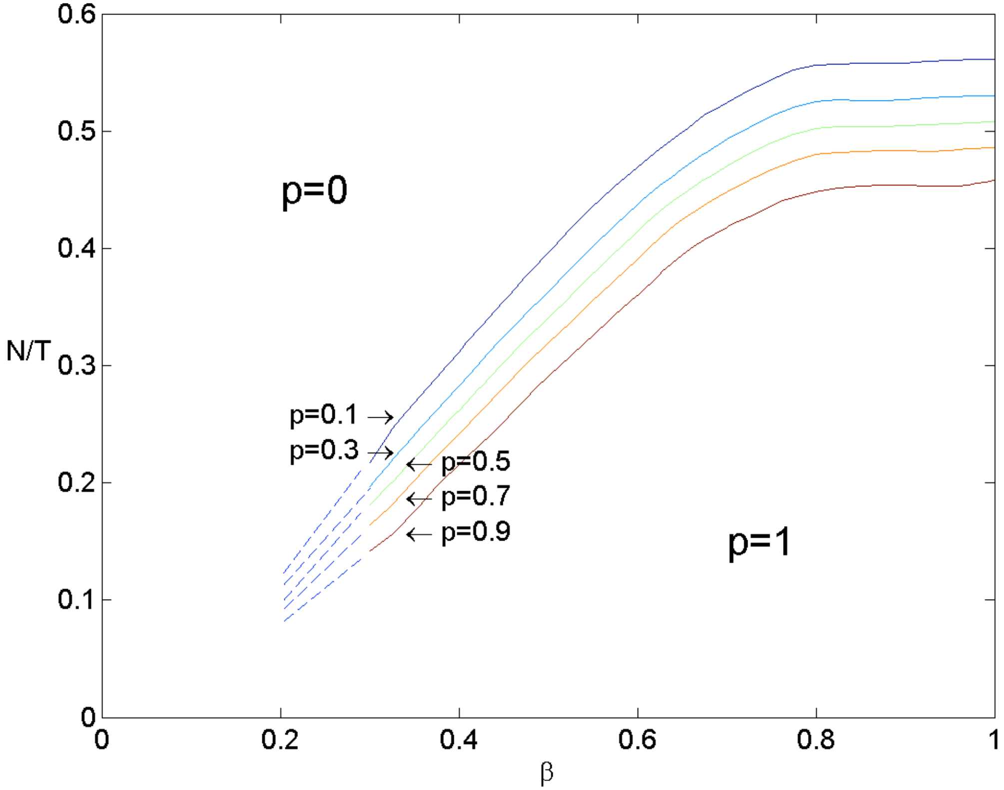

N=2; T=8
Risk
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
2
1
0
-1
-2
w2
-2
-1
0
1
2
w1

Feasibility of optimization under ES

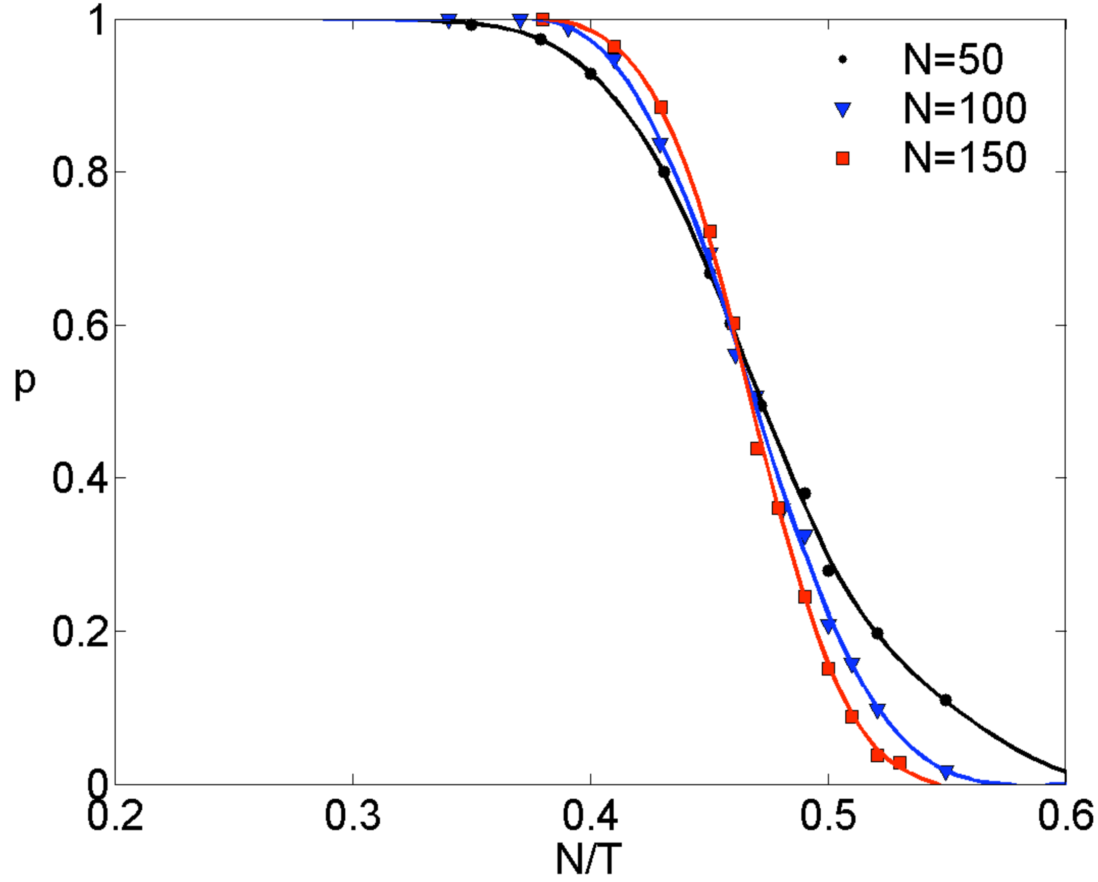


Probability of the existence of an optimum under CVaR.
F is the standard normal distribution. Note the scaling in $N/\sqrt{T}$.

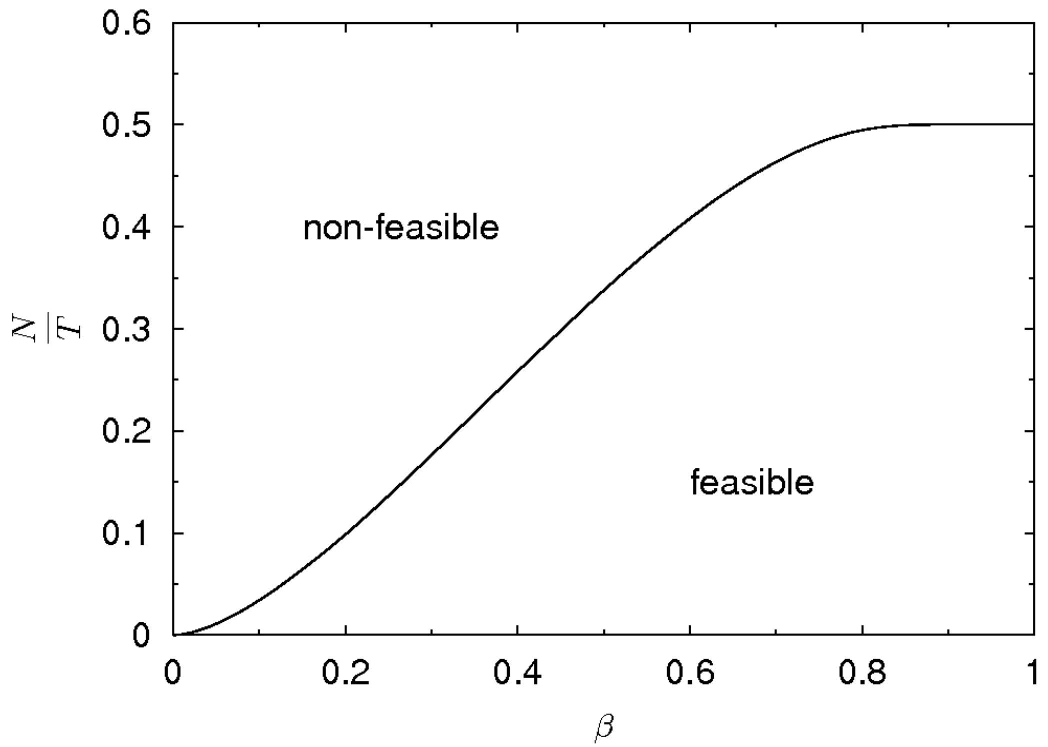
For ES the critical value of *N/T* depends on the threshold β



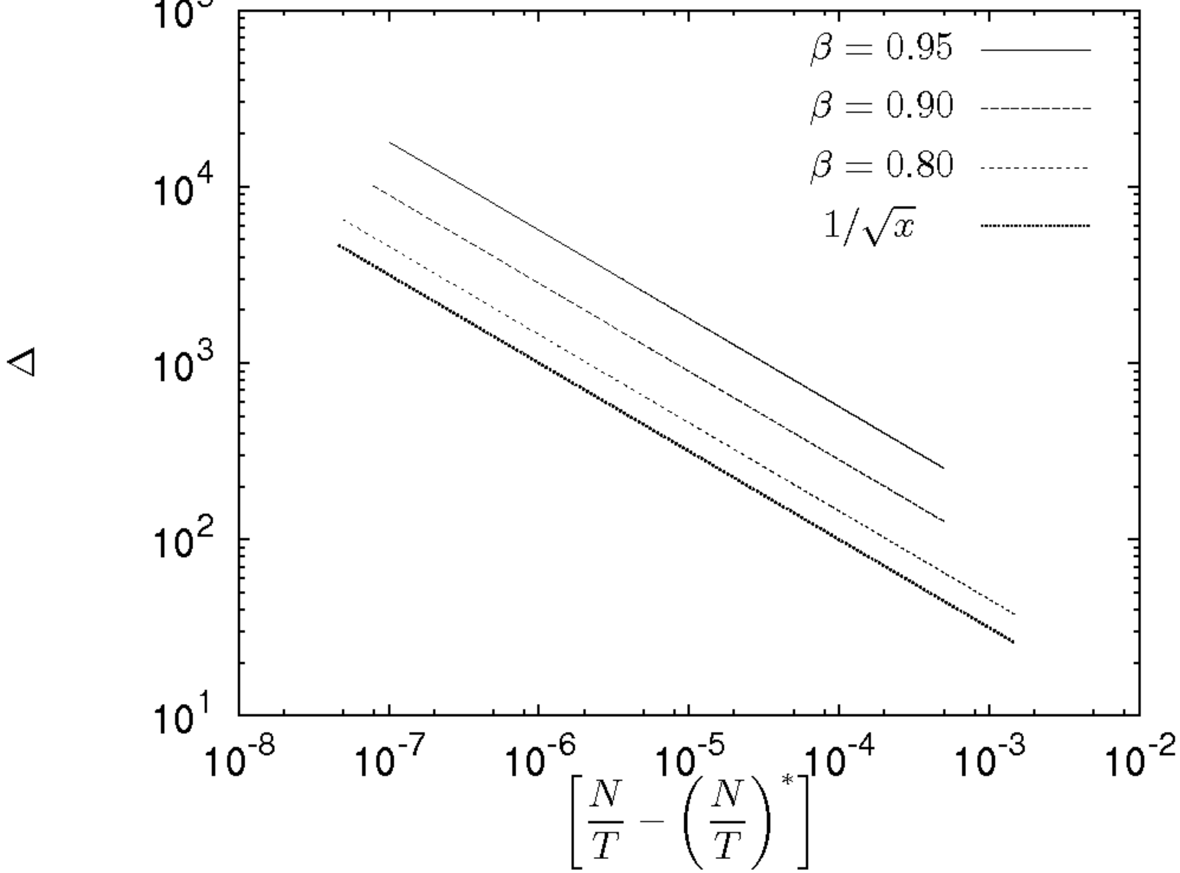
With increasing *N*, *T* (*N/T*= fixed) the transition becomes sharper and sharper…



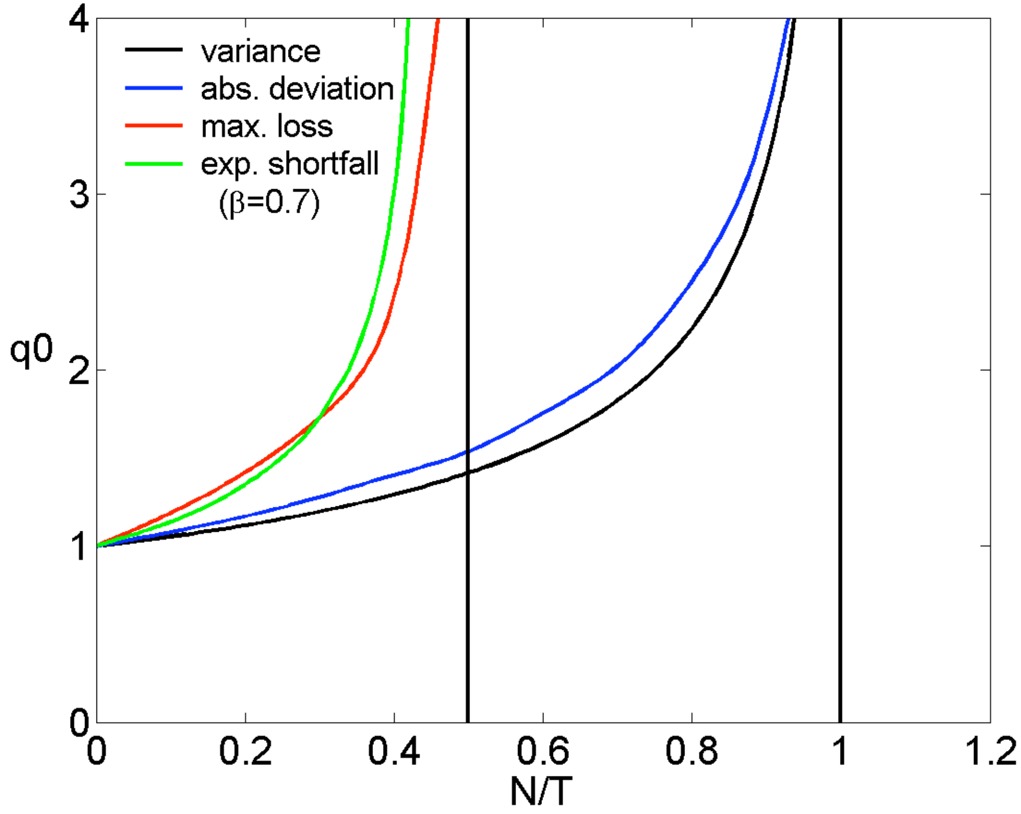
…until in the limit N, $T \rightarrow \infty$ with N/T= fixed we get a „phase boundary”. The exact phase boundary has since been obtained by Ciliberti, Kondor and Mézard from replica theory.



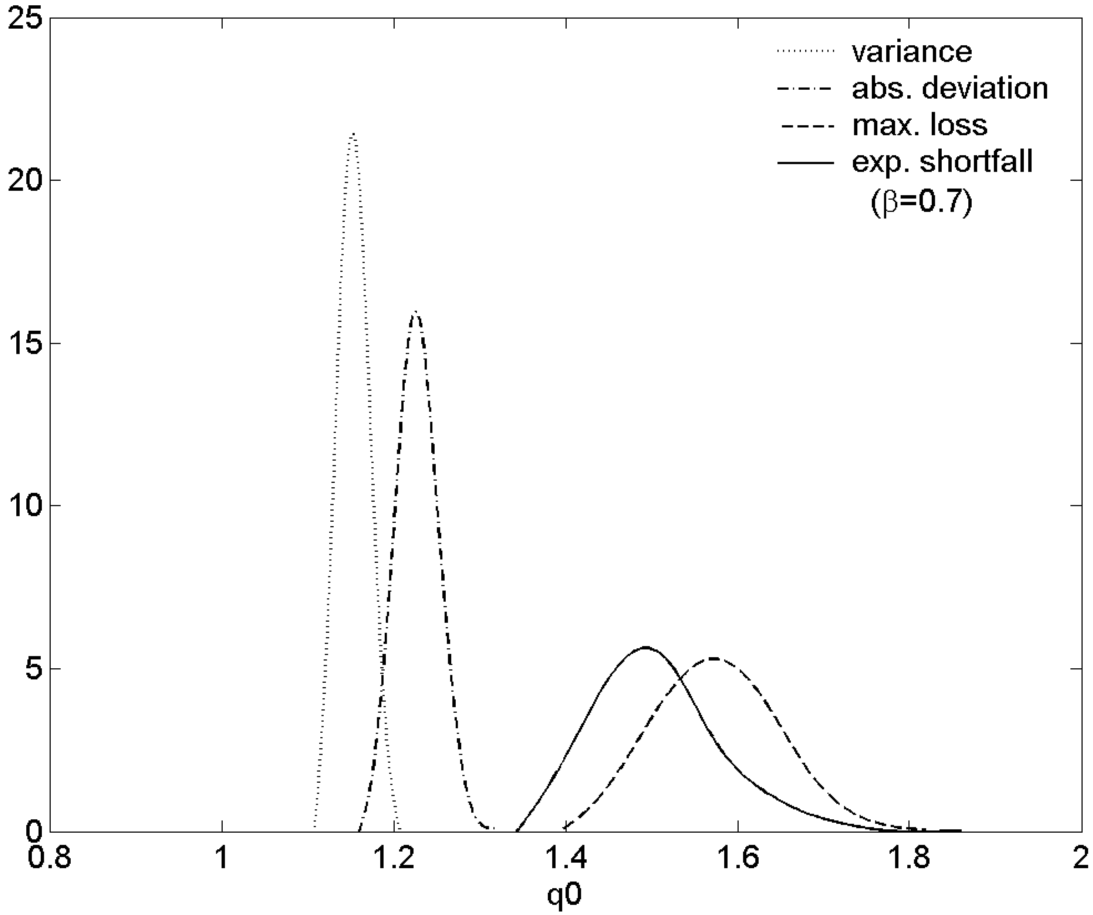
Scaling: same exponent



The mean relative error in portfolios optimized under various risk measures blows up as we approach the phase boundary



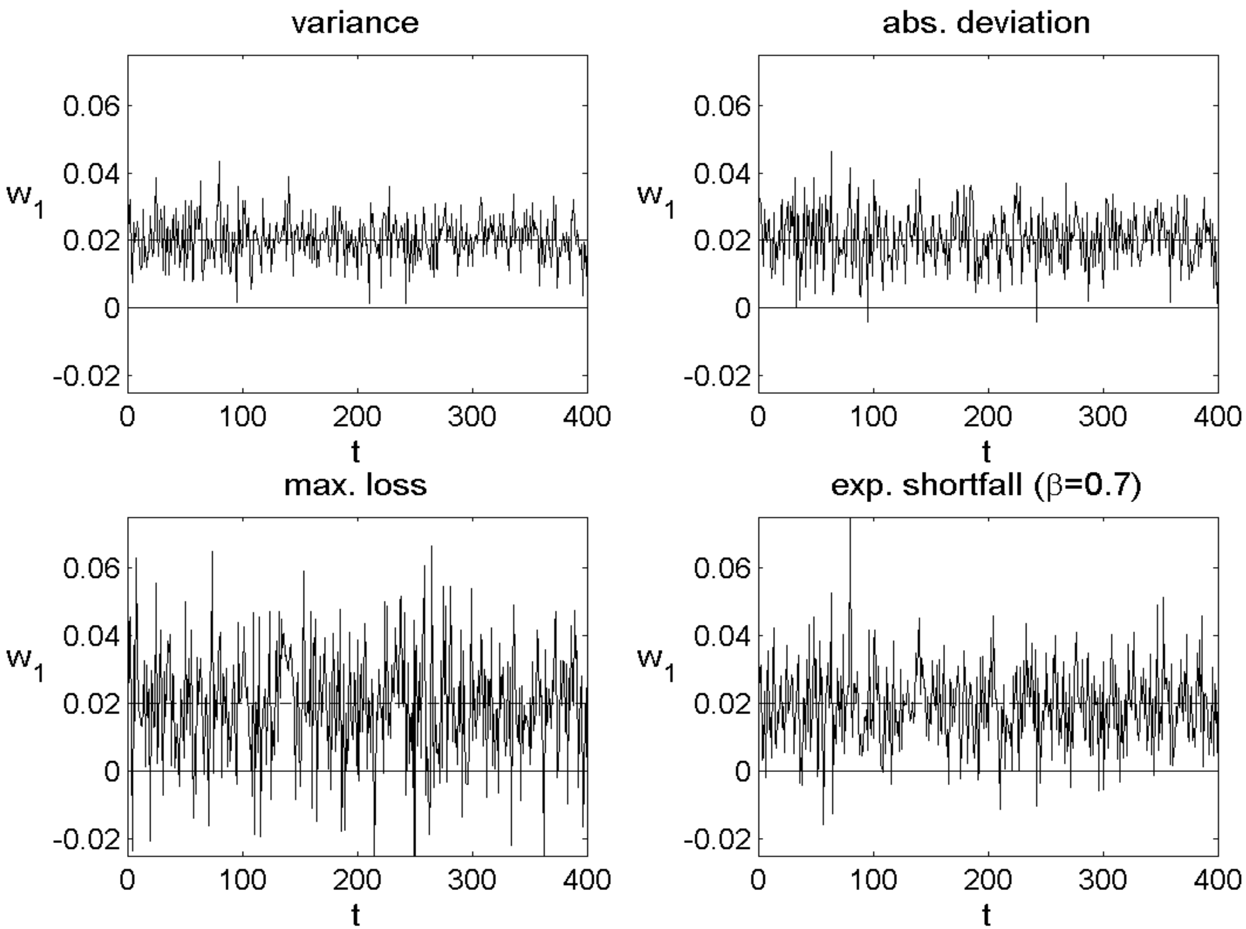
Distributions of q_o for various risk measures



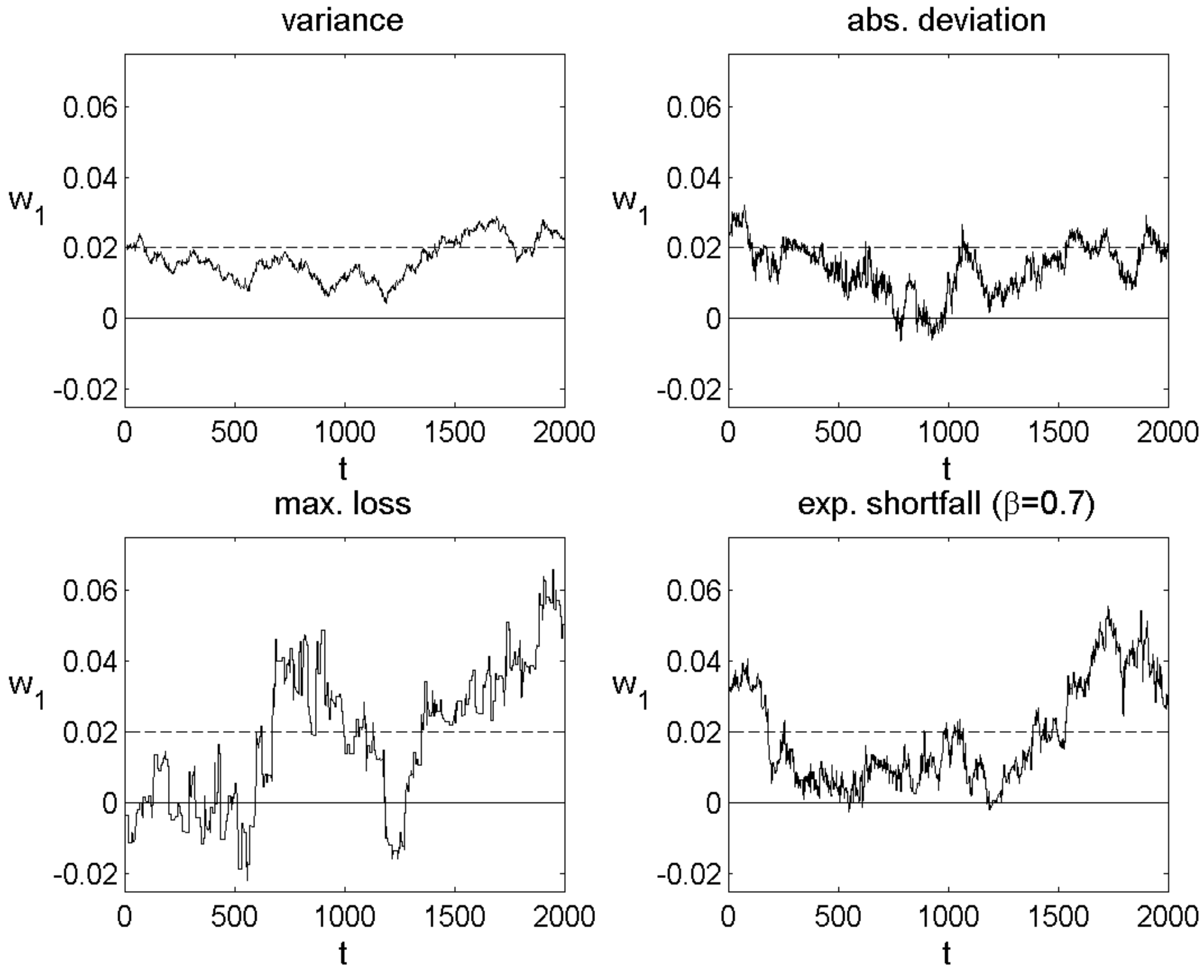
Instability of portfolio weights

Similar trends can be observed if we look into the weights of the optimal portfolio: the weights display a high degree of instability already for variance optimized portfolios, but this instability is even stronger for mean absolute deviation, expected shortfall, and maximal loss.

Instability of weights for various risk measures, non-overlapping windows



Instability of weights for various risk measures, overlapping weights



A wider context

- The critical phenomena we observe in portfolio selection are analogous to the phase transitions discovered recently in some hard computational problems, they represent a new „random Gaussian" universality class within this family, where a number of modes go soft in rapid succession, as one approaches the critical point.
- Filtering corresponds to discarding these soft modes.

A wider context

- The critical phenomena we observe in portfolio selection are analogous to the phase transitions discovered recently in some hard computational problems, they represent a new „random Gaussian" universality class within this family, where a number of modes go soft in rapid succession, as one approaches the critical point.
- Filtering corresponds to discarding these soft modes.

A prophetic quotation:

P.W. Anderson: The fact is that the techniques which were developed for this apparently very specialized problem of a rather restricted class of special phase transitions and their behavior in a restricted region are turning out to be something which is likely to spread over not just the whole of physics but the whole of science.

In a similar spirit...

- I think the phenomenon treated here, that is the sampling error catastrophe due to lack of sufficient information appears in a much wider set of problems than just the problem of investment decisions. (E.g. multivariate regression, all sorts of linearly programmable technology and economy related optimization problems, microarrays, etc.)
- Whenever a phenomenon is influenced by a large number of factors, but we have a limited amount of information about this dependence, we have to expect that the estimation error will diverge and fluctuations over the samples will be huge.

The appearence of powerful tools from statistical physics (random matrices, phase transition concepts, scaling, universality, etc. and replicas) is an important development that enriches finance theory

Summary

- If we do not have sufficient information we cannot make an intelligent decision – so far this is a triviality
- The important message here is that there is a critical point where the error diverges, and its behaviour is subject to universal scaling laws

Appendix I: Optimization and statistical mechanics

- Any convex optimization problem can be transformed into a problem in statistical mechanics, by promoting the objective function into a Hamiltonian, and introducing a fictitious temperature. At the end we can recover the original problem in the limit of zero temperature.
- Averaging over the time series segments (samples) is similar to what is called quenched averaging in the statistical physics of random systems: one has to average the logarithm of the partition function (i.e. the cumulant generating function).
- Averaging can then be performed by the replica trick – a heuristic, but very powerful method that is on its way to become firmly established by mathematicians (Guerra and Talagrand).

The first application of replicas in a finance context: the ES phase boundary (A. Ciliberti, I.K., M. Mézard)

ES is the average loss above a high threshold β (a conditional expectation value). Very popular among academics and slowly spreading in practice. In addition, as shown by Uryasev and Rockafellar, the optimization of ES can be reduced to linear programming, for which very fast algorithms exist.

Portfolios optimized under ES are much more noisy than those optimized under either the variance or absolute deviation. The critical point of ES is always below $N/T = 1/2$ and it depends on the threshold, so it defines a phase boundary on the N/T- β plane.

The measure ES can become unbounded from below with a certain probability for any finite N and T , and then the optimization is not feasible!

The transition for finite N,T is smooth, for $N,T \to \infty$ it becomes a sharp phase boundary that separates the region where the optimization is feasible from that where it is not.

Formulation of the problem

- The time series of returns $x_{i\tau}$, $i = 1, \ldots N$, $\tau = 1, \ldots T$,
- The objective functi $E_\beta[v, \{w_i\}, \{u_\tau\}; \{x_{i\tau}\}] = (1-\beta)Tv + \sum_{t=\tau}^{T} u_\tau$
- The variables: $\mathbf{Y} \equiv \{w_1, \ldots w_N\, u_1, \ldots u_T\, v\}$
- The linear programming problem:

$$u_\tau \geq 0\ , \quad u_\tau + v + \sum_{i=1}^{N} x_{i\tau} w_i \geq 0 \quad \forall \tau$$

- Normali: $\sum_{i=1}^{N} w_i = N$ $\qquad w_i \geq -W \qquad W \to \infty$

Associated statistical mechanics problem

- Partition functio $Z_\gamma[\{x_{it}\}] = \int_V d\mathbf{Y} \exp\Big[-\gamma E_\beta[\mathbf{Y};\{x_{it}\}]\Big]$

- Free energy: $\lim_{\gamma\to\infty} \frac{-1}{N\gamma} \log Z_\gamma[\{x_{it}\}]$

$$\epsilon[\{x_{it}\}] = \lim_{N\to\infty} \frac{\min E[\{x_{it}\}]}{N} = \lim_{N\to\infty} \lim_{\gamma\to\infty} \frac{-1}{N\gamma} \log Z_\gamma[\{x_{it}\}]$$

The partition function

$$Z_\gamma[\{x_{it}\}] = \int_{-\infty}^{+\infty} dv \int_0^{+\infty} \prod_{t=1}^{T} du_t \int_{-\infty}^{+\infty} \prod_{i=1}^{N} dw_i \int_{-i\infty}^{+i\infty} d\lambda \exp\left[\lambda\left(\sum_{i=1}^{N} w_i - N\right)\right] \times$$

$$\times \int_0^{+\infty} \prod_{t=1}^{T} d\mu_t \int_{-i\infty}^{+i\infty} \prod_{t=1}^{T} d\hat{\mu}_t \exp\left[\sum_{t=1}^{T} \hat{\mu}_t \left(u_t + v + \sum_{i=1}^{N} x_{it} w_i - \mu_t\right)\right] \exp\left[-\gamma(1-\beta)Tv - \gamma \sum_{t=1}^{T} u_t\right]$$

Lagrange multipliers: $\lambda, \mu, \hat{\mu}$

Replicas

- Trivial identity $$\overline{\log Z} = \lim_{n\to 0} \frac{\partial \overline{Z^n}}{\partial n}$$

- We consider *n* identical replicas: $\mathbf{Y^1}, ..., \mathbf{Y^n}$

- The probability distribution of the *n*-fold replicated system:

$$P_\gamma(\mathbf{Y^1}, ..., \mathbf{Y^n}) = \frac{1}{Z_\gamma^n[\{x_{it}\}]} exp\Big[-\gamma \sum_{a=1}^{n} E_\beta[\mathbf{Y^a}; \{x_{it}\}] \Big]$$

- At an appropriate moment we have to analytically continue to real n's

Averaging over the random samples

$$\overline{Z_\gamma^n[\{x_{it}\}]} \sim \int_{-\infty}^{+\infty} \prod_{a=1}^{n} dv^a \int_{-\infty}^{+\infty} \prod_{a,b} dQ^{ab} \int_{-i\infty}^{+i\infty} \prod_{a,b} d\hat{Q}^{ab} \exp\left\{ N \sum_{a,b} Q^{ab}\hat{Q}^{ab} - N \sum_{a,b} \hat{Q}^{ab} - \gamma(1-\beta)T \sum_a v^a \right.$$

$$\left. -Tn\log\gamma + T\log\hat{Z}_\gamma\left(\{v^a\},\{Q^{ab}\}\right) - \frac{T}{2}\mathrm{Tr}\log Q - \frac{N}{2}\mathrm{Tr}\log\hat{Q} - \frac{nN}{2}\log 2 \right\}$$

$$\hat{Z}_\gamma\left(\{v^a\},\{Q^{ab}\}\right) \equiv \int_{-\infty}^{+\infty} \prod_{a=1}^{n} dy^a \exp\left[-\frac{1}{2}\sum_{a,b=1}^{n} (Q^{-1})^{ab}(y^a - v^a)(y^b - v^b) + \gamma \sum_{a=1}^{n} y^a \theta(-y^a) \right]$$

$$Q^{ab} = \frac{1}{N}\sum_{i=1}^{N} w_i^a w_i^b \;, \quad a,b = 1,\ldots n$$

Replica-symmetric Ansatz

- By symmetry considerations:

$$Q^{ab} = \begin{cases} q_1 & \text{if } a = b \\ q_0 & \text{if } a \neq b \end{cases} \; ; \quad \hat{Q}^{ab} = \begin{cases} \hat{q}_1 & \text{if } a = b \\ \hat{q}_0 & \text{if } a \neq b \end{cases}$$

- Saddle point condition:

$$\epsilon(v, q_0, \Delta) = \frac{1}{2\Delta} + \Delta \left[t(1-\beta)v - \frac{q_0}{2} + \frac{t}{2\sqrt{\pi}} \int_{-\infty}^{+\infty} dse^{-s^2} g(v + s\sqrt{2q_0}) \right]$$

$$g(x) = \begin{cases} 0 & x \geq 0 \,, \\ x^2 & -1 \leq x < 0 \,, \\ -2x - 1 & x < -1 \,. \end{cases}$$

$$\begin{aligned} 1 - \beta + \frac{1}{2\sqrt{\pi}} \int dse^{-s^2} g'(v + s\sqrt{2q_0}) &= 0 \,, \\ -1 + \frac{t}{\sqrt{2\pi q_0}} \int dse^{-s^2} s \; g'(v + s\sqrt{2q_0}) &= 0 \,. \end{aligned}$$

Condition for the existence of a solution to the linear programming problem

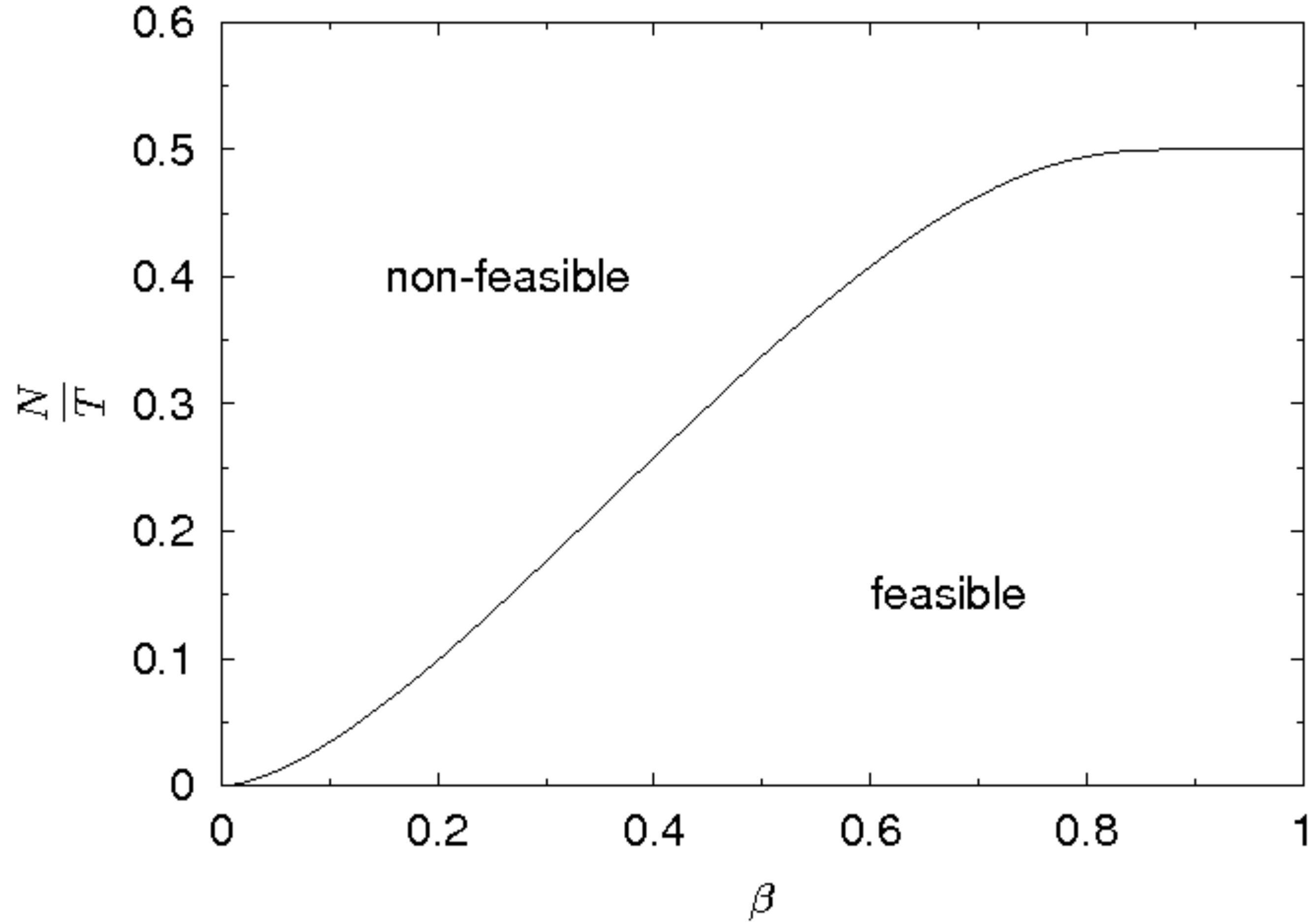
The meaning of the parameter Δ :

$$\Delta/\gamma \sim \Delta q = (q_1 - q_0) = \frac{1}{N}\sum_{i=1}^{N}\left(w_i^{(1)}\right)^2 - \frac{1}{N}\sum_{i=1}^{N} w_i^{(1)} w_i^{(2)} \sim \overline{w^2} - \overline{w}^2$$

Equation of the phase boundary:

$$t(1-\beta)v - \frac{q_0}{2} + \frac{t}{2\sqrt{\pi}}\int_{-\infty}^{+\infty} dse^{-s^2} g(v + s\sqrt{2q_0}) \geq 0 \qquad T \equiv 1/t$$

$$\beta \longrightarrow 1 \qquad \frac{1}{t^*} = \frac{1}{2} - \mathcal{O}\left[(1-\beta)^3 e^{-\left(4\pi(1-\beta)^2\right)^{-1}}\right]$$


0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
N/T
β
non-feasible
feasible

Appendix II: Portfolio optimization and linear regression

Portfolios:

$$\varsigma_p = -\mu + \sum_{i=1}^{N} w_i r_i \qquad \sum_{i=1}^{N} w_i = 1$$

$$\min_{\mu, w_i} \mathrm{Var}(\varsigma_p)$$

$$\mathrm{Var}(\varsigma_p) = \mathrm{Var}(r_p) = \sum_{i,j=1}^{N} C_{ij} w_i w_j$$

Linear regression:

- $$y = \beta_0 + \sum_{i=1}^{N-1} \beta_i x_i + \varepsilon \qquad \min_{\beta_0, \beta_1, \ldots, \beta_{N-1}} \mathrm{Var}(\varepsilon)$$

$$R(\beta) = \mathrm{Var}(\varepsilon) = E(\varepsilon^2) = E\left(y - \beta_0 - \sum_{i=1}^{N-1} \beta_i x_i\right)^2$$

$$\frac{\partial R(\beta)}{\partial \beta_j} = -2\left[\sum_{i=1}^{N-1} \mathrm{Cov}(x_j, x_i)\beta_i - \mathrm{Cov}(x_j, y)\right] = 0$$

$$\frac{\partial R(\beta)}{\partial \beta_0} = 2\left[\beta_0 + \sum_{i=1}^{N-1} E(x_i) - E(y)\right] = 0$$

Equivalence of the two

$$-\varepsilon = \beta_0 - y + \sum_{i=1}^{N-1} \beta_i x_i$$

$$-\varepsilon = \beta_0 - y + y\sum_{i=1}^{N-1} \beta_i - y\sum_{i=1}^{N-1} \beta_i + \sum_{i=1}^{N-1} \beta_i x_i =$$

$$= \beta_0 - y\left(1 - y\sum_{i=1}^{N-1} \beta_i\right) + \sum_{i=1}^{N-1} \beta_i (x_i - y)$$

Translation

$$r_i = x_i - y \qquad r_N = -y$$

$$\mu_p = -\beta_0 \qquad w_i = \beta_i$$

$$w_N = 1 - \sum_{i=1}^{N-1} \beta_i \qquad \varsigma_p = -\varepsilon$$

Minimizing the residual error for and infinitely large sample

$$\varepsilon = y - \beta_0 - \sum_{i=1}^{N-1} \beta_i x_i \qquad E(\varepsilon) = 0$$

$$\min_{\beta_0, \beta_{1,\ldots,}\beta_{N-1}} \mathrm{Var}(\varepsilon)$$

$$\beta_0^*, \beta_1^*, \ldots, \beta_{N-1}^*$$

$$\varepsilon^* = y - \beta_0^* - \sum_{i=1}^{N-1} \beta_i^* x_i \qquad \sigma_\varepsilon^* = \sqrt{\mathrm{Var}(\varepsilon^*)}$$

Minimizing the residual error for a sample of length T

$$\varepsilon_t = y_t - \beta_0 - \sum_{i=1}^{N-1} \beta_i x_{it} \qquad \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t = 0$$

$$\min_{\beta_0, \beta_{1,\dots,}\beta_{N-1}} \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2$$

$$\hat{\beta}_0^*, \hat{\beta}_1^*, \dots, \hat{\beta}_{N-1}^* \qquad \hat{\varepsilon}^* = y - \hat{\beta}_0^* - \sum_{i=1}^{N-1} \hat{\beta}_i^* x_i \qquad \hat{\sigma}_\varepsilon^* = \sqrt{\mathrm{Var}\left(\hat{\varepsilon}^*\right)} > \sigma_\varepsilon^*$$

The relative error

$$q_0 = \frac{\hat{\sigma}_\varepsilon}{\sigma_\varepsilon} = \sqrt{\frac{\mathrm{Var}\left(\varepsilon^*\right)}{\mathrm{Var}\left(\hat{\varepsilon}^*\right)}} > 1$$