The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2229-8

School and Workshop on Market Microstructure: Design, Efficiency and Statistical Regularities

21 - 25 March 2011

Regularizing Portfolio Optimization

Imre KONDOR
Eotvos Lorand University
Faculty of Science
Institute of Theoretical Physics, HU-1117
Budapest
HUNGARY

REGULARIZING PORTFOLIO OPTIMIZATION

Imre Kondor

Eötvös University, Budapest

School and Workshop on Market Microstructure: Design, Efficiency and Statistical Regularities

ICTP, Trieste, 21-25 March, 2011

This work has been supported by the National Office for Research and Technology under grant No. KCKHA005

Coworkers

- **István Varga-Haszonits** (ELTE PhD student →MSCI, Budapest)
- **Susanne Still** (University of Hawaii)
- **Matteo Marsili** (ICTP, Trieste)
- **Fabio Caccioli** (PhD student, SISSA, → Santa Fe Institute)

Contents

- Sensitivity of risk measures to estimation error
- Regularization

ESTIMATION ERROR AND INSTABILITY IN PORTFOLIO SELECTION

The problem

- Institutional portfolios are large, number of items $N >> 1$
- Sampling frequency cannot be high (one doesn't rebalance a portfolio every second), the look-back period cannot be long (lack of stationarity), therefore the size of the samples (length of time series T) is not large enough.
- One cannot make a good decision without sufficient information

Estimation error

- The estimation error strongly depends on the ratio N/T: for $N/T \to 0$ the error is small, for $N/T \sim O(1)$ it can be very large, at a critical value $(N/T)_c$ it actually diverges. (An algorithmic phase transition takes place.)
- Large portfolios are close to this critical point, or they are even on the „wrong side" of it, with $N/T > 1$.
- This is a generic problem in high dimensional statistical optimization (social, medical sciences, gene chips, climate research, etc.)

Classical portfolio theory

- Markowitz: the risk measure is the variance to be minimized subject to the budget constraint and possible other linear constraints.
- Its critical point is at $(N/T)_c = 1$, where the covariance matrix first develops a zero eigenvalue.
- The variance is not an adequate risk measure for fat-tailed items.
- A host of alternative risk measures are being considered: mean absolute deviation, VaR, Expected Shortfall, other coherent measures, etc.

Consider the following trivial investment problem $(N = 2,\ T = 1)$

$N = 2$ assets with returns: x_1 and x_2, iid normal, say, and a sample of size $T = 1$, that is a *single* observation. Let us choose the Maximal Loss (ML), the best combination of the worst losses, as our risk measure:

$$ML = \min_w \max_t \left\{ -\sum w_i x_{it} \right\}$$

This is a coherent measure, in the sense of Artzner at al., a limiting case of Expected Shortfall (ES).

In the particular case of $N=2$, $T=1$

$$ML = \min_w \{-(w_1 x_1 + w_2 x_2)\}$$

subject to $w_1 + w_2 = 1$, that is $w_1 = 1 - w_2$

Our optimization problem is then

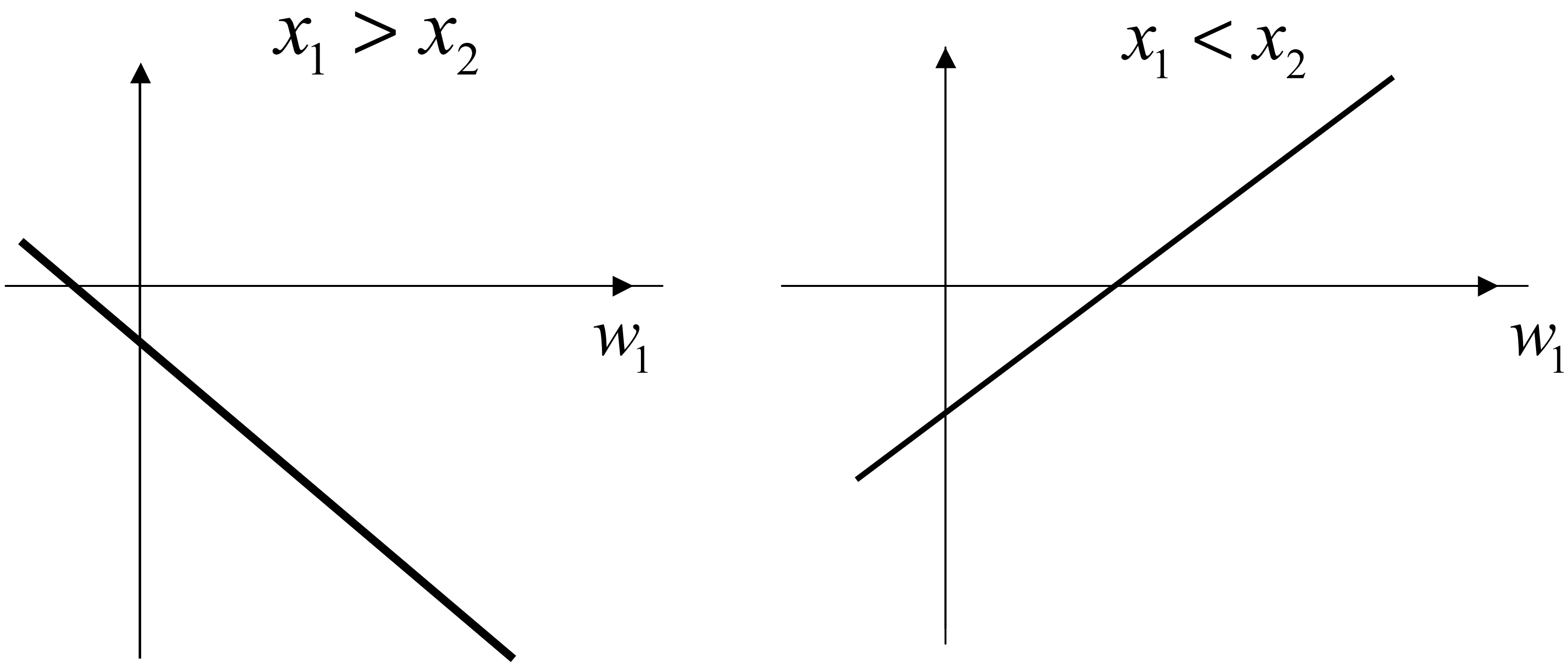
$$ML = \min_w \{-x_2 - (x_1 - x_2) w_1\}$$

Obviously, the solution is

$w_1^* = \infty$, $w_2^* = -\infty$, for $x_1 > x_2$ and

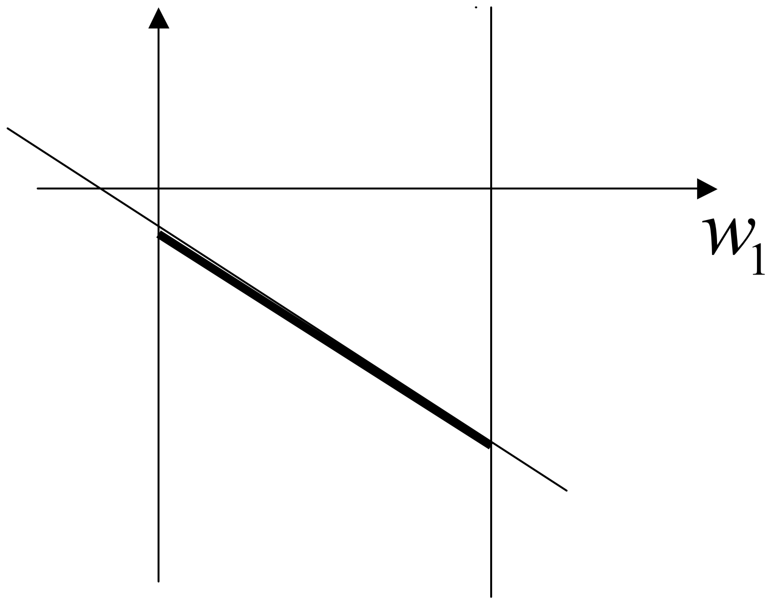
$w_1^* = -\infty, w_2^* = \infty$, for $x_1 < x_2$.

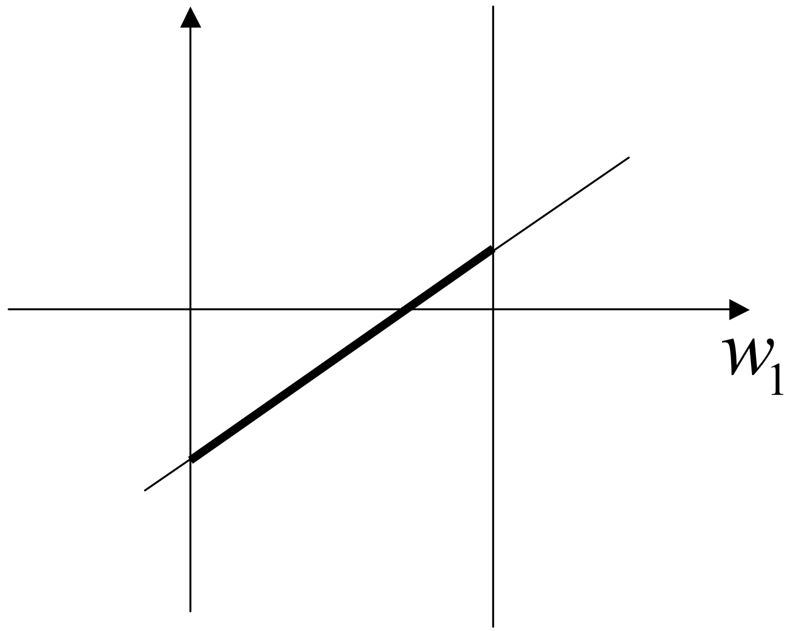
The two cases



ML as a risk measure is unbounded with probability 1, if $N = 2$ and $T = 1$.

If there are some constraints, e.g. short selling is banned, $w_i > 0$





$w_1^* = 1 \quad w_2^* = 0$

$w_1^* = 0 \quad w_2^* = 1$

Prob. 1/2

Prob. 1/2

So, for $N = 2$, $T = 1$ (and also for any $N › T$)

Without constraint:

the risk measure ML is not bounded with probability 1, there is no solution, we are tempted to go infinitely long in the dominant item, and infinitely short in the dominated one.

With constraint:

the risk measure is bounded but monotonic, so with probability 1 we go as long as allowed by the constraint in the dominating item, and take a zero position in the dominated one.

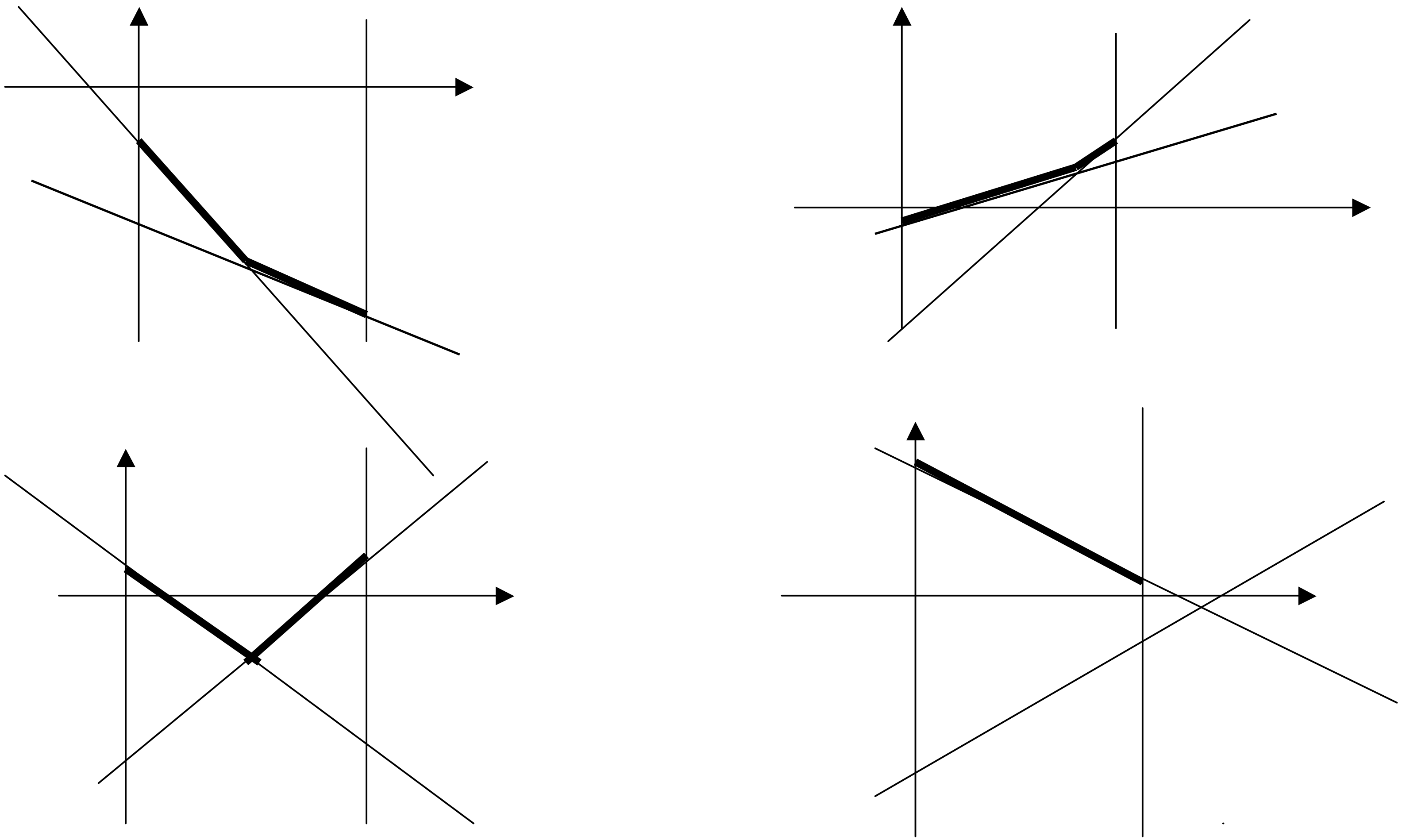
The same for $N = 2$ and $T = 2$

$ML = \min_w \max_y \{y_1, y_2\}$, where $y_1 = -x_{21} - (x_{11} - x_{21})w_1$ and $y_2 = -x_{22} - (x_{12} - x_{22})w_1$

There is no solution if $x_{11} > x_{21}$ *and* $x_{12} > x_{22}$, *or* $x_{11} < x_{21}$ *and* $x_{12} < x_{22}$, that is when one of the items dominates the other in the sample. This happens with probability ½ (assuming iid variables, say).

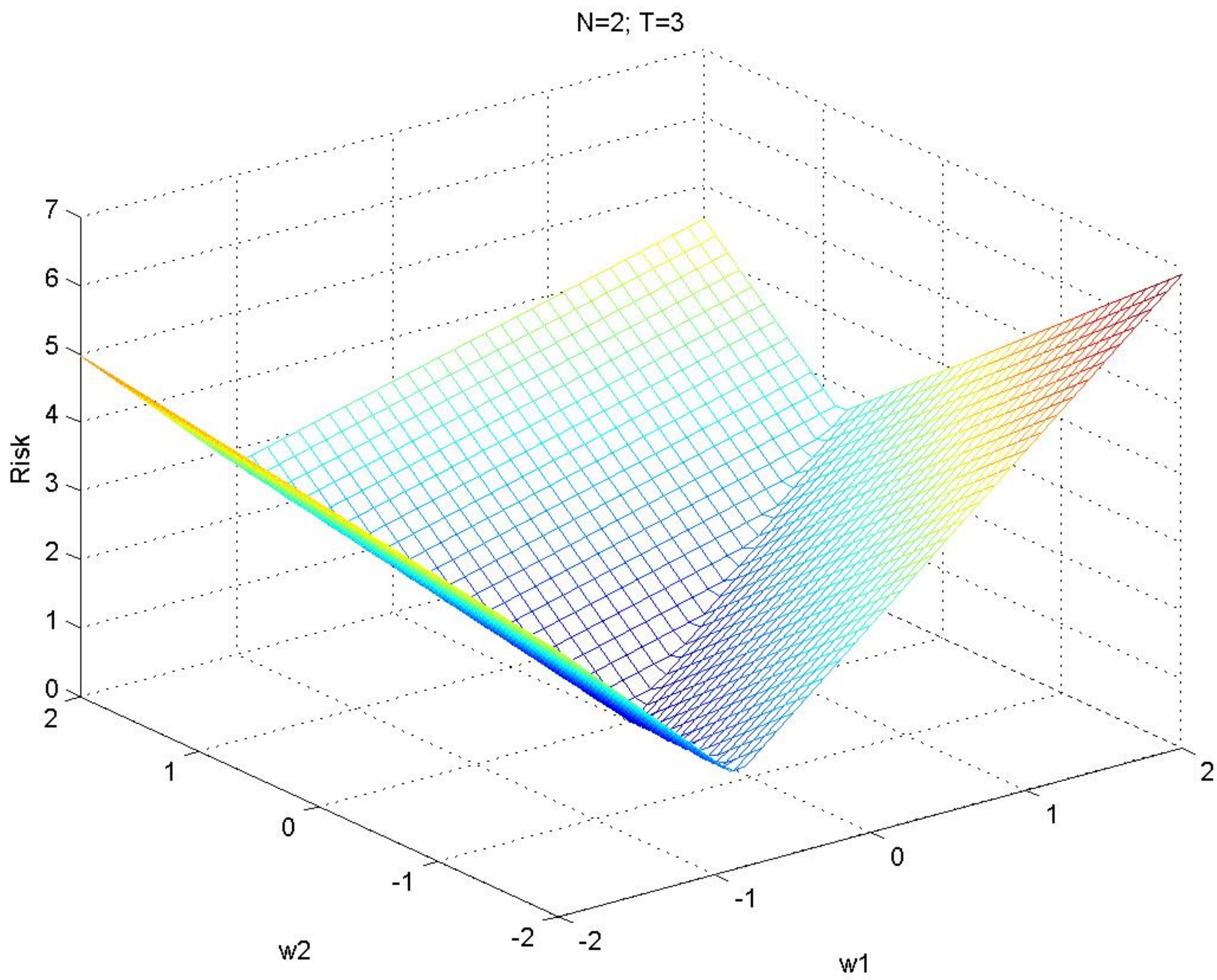
There is a finite solution if $x_{11} > x_{21}$ *and* $x_{12} < x_{22}$, *or* $x_{11} < x_{21}$ *and* $x_{12} > x_{22}$, that is when none of the items dominates the other. The probability of this event is 1/2 again.

Geometrically

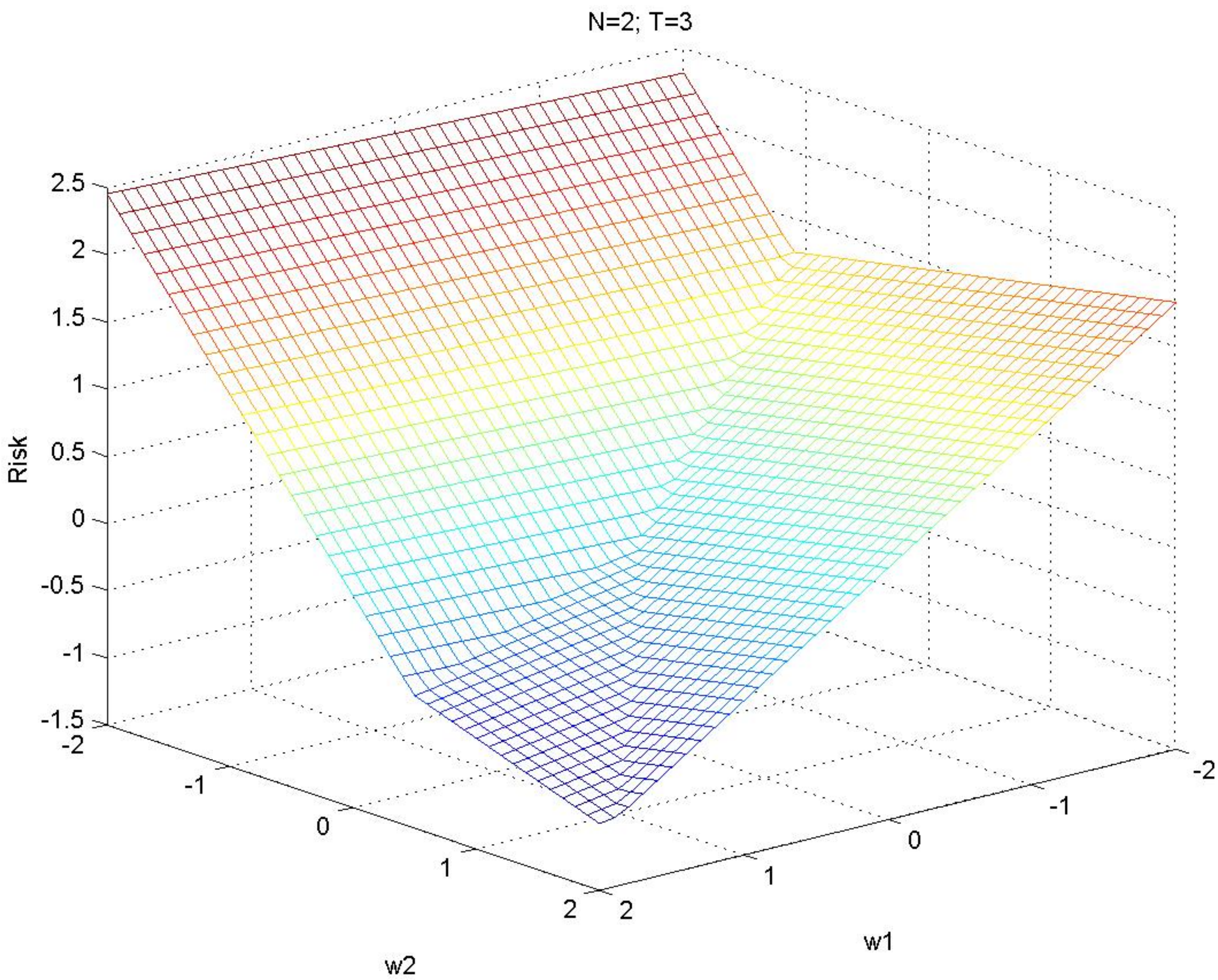


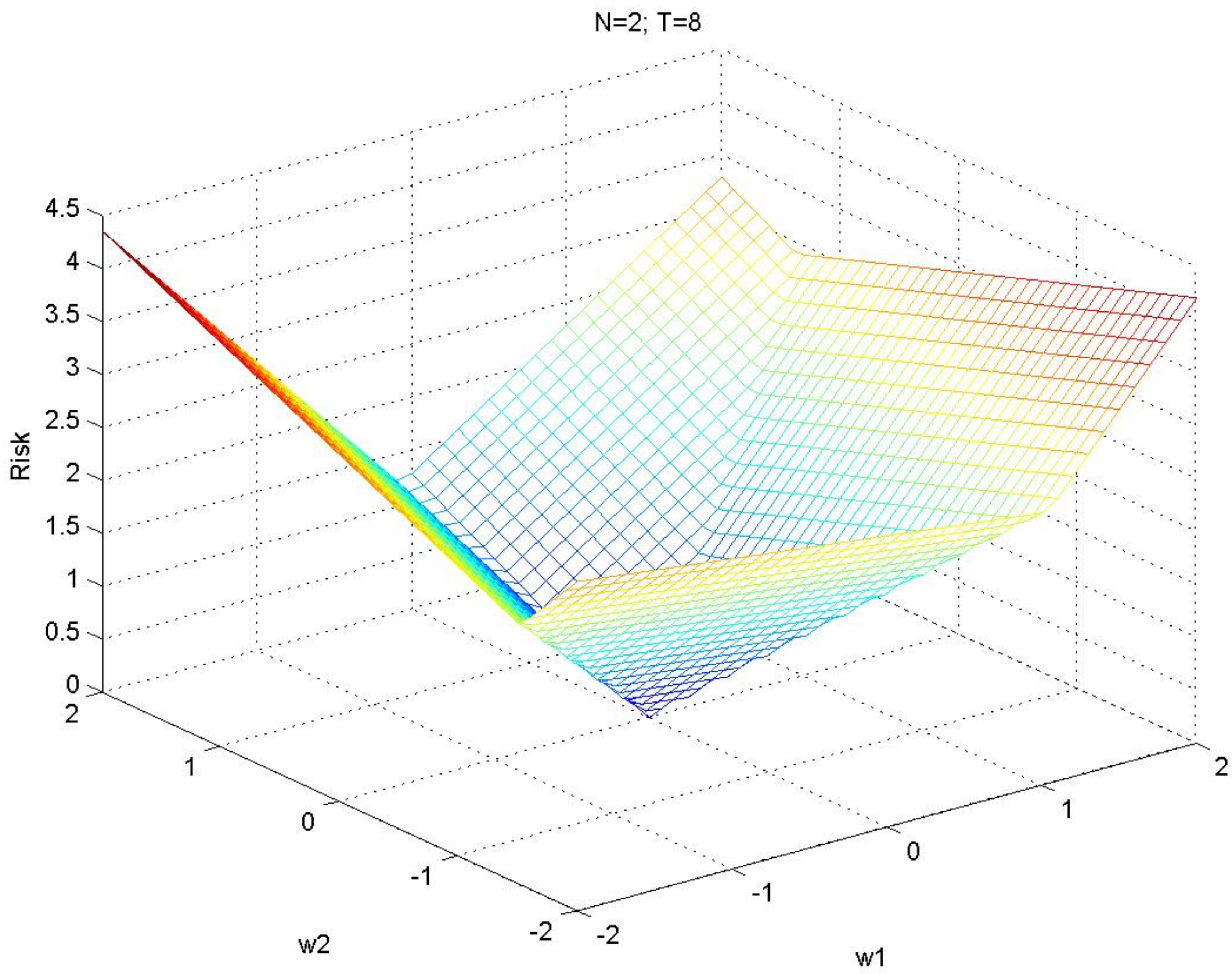
For N =2, T =2 there is no solution with probability 1/2, and there is a finite solution with probability 1/2. When one of the items dominates, there is no finite solution, unless we impose some constraints. Then we go as long as allowed by the constraints in the dominating item, and take zero position as in the dominated one.

When neither of them dominates, we have a finite solution that may or may not fall inside the allowed region.



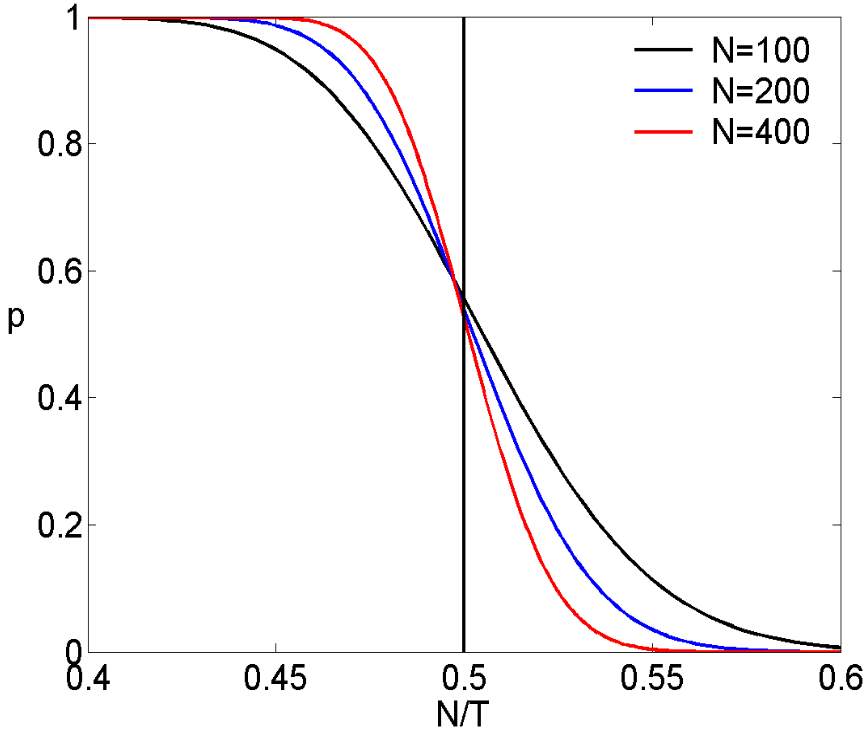
N=2; T=3
Risk
7
6
5
4
3
2
1
0
w2
2
1
0
-1
-2
w1
-2
-1
0
1
2


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

- The existence of a finite solution depends on the sample, therefore it is a probabilistic event.
- Although the constraints may prevent the solution from running away to infinity, they do not quite stabilize it: If a set of weights vanishes for a given sample, a different set will vanish for the next sample, the solution jumps around on the boundaries of the allowed region.
- The smaller the ratio N/T, the larger the probability of a finite solution, and the smaller the generalization error.
- In real life N/T is almost never small; the limit $N, T \rightarrow \infty$, with N/T = fixed, is closer to reality.

Probability of finding a solution for the minimax problem (general N and T, elliptic underlying distribution):



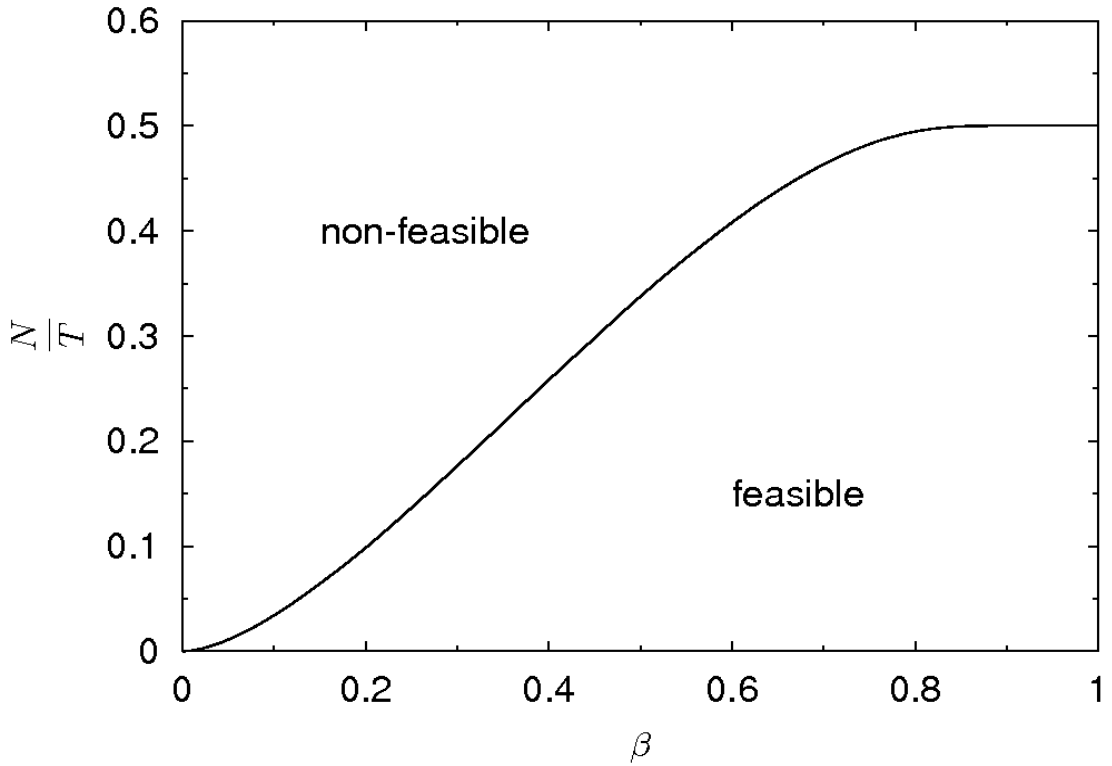
$$p = \frac{1}{2^{T-1}} \sum_{k=N-1}^{T-1} \binom{T-1}{k}$$

In the limit $N, T \to \infty$, with N/T fixed, the transition becomes sharp at $N/T = ½$. The estimation error diverges as we go to $N/T = ½$ from below.

Generalization: Expected Shortfall

ES is the conditional expectation value of losses above a high threshold. It has an obvious meaning, it is easy to determine form historical time series, and can be optimized via linear programming. ML is the limiting case of ES, when the threshold goes to 1. ES shows the same instability as ML, but the locus of this instability depends not only on *N/T*, but also on the threshold β above which the conditional average is calculated. So there will be a critical line.

This critical line or phase boundary for ES has been obtained numerically by I. K., Sz. Pafka, G. Nagy: Noise sensitivity of portfolio selection under various risk measures, *Journal of Banking and Finance*, **31**, 1545-1573 (2007) and calculated analytically in A. Ciliberti, I. K., and M. Mézard: On the Feasibility of Portfolio Optimization under Expected Shortfall, *Quantitative Finance*, **7**, 389-396 (2007)



The estimation error diverges as one approaches the phase boundary from below

Generalization stage II: Coherent measures

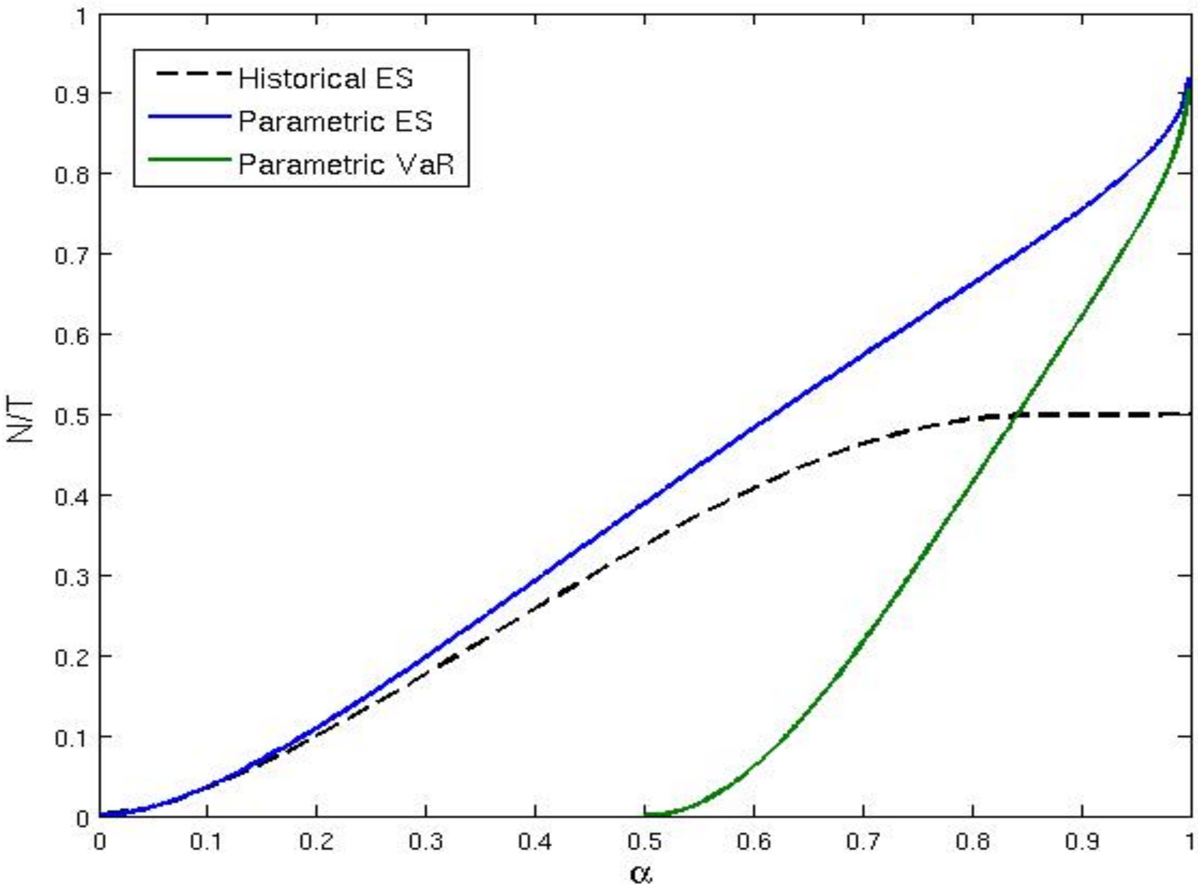
- The intuitive explanation for the instability of ES and ML is that for a *given finite sample* there may exist a dominant item (or a dominant combination of items) that produces a larger return at each time point than any of the others, *even if no such dominance relationship exist between them on very large samples*. This mirage of arbitrage leads the investor to believe that if she goes extremely long in the dominant item and extremely short in the rest, she can produce an arbitrarily large return on the portfolio, at a risk that goes to minus infinity (i.e. no risk).
- The same consideration is true for any coherent risk measure: I. Kondor and I. Varga-Haszonits: Instability of portfolio optimization under coherent risk measures, Advances in Complex Systems, **13,** 425-437 (2010)

Further generalization

- As a matter of fact, this type of instability appears even beyond the set of coherent risk measures, and may appear in downside risk measures in general.
- By far the most widely used risk measure today is Value at Risk (VaR). It is a downside measure. It is not convex, therefore the stability problem of its historical estimator is ill-posed.
- Parametric VaR, however, is convex, and this allows us to study the stability problem. Along with VaR, we also look into the closely related parametric estimate for ES.
- Parametric estimates are expected to be more stable than historical ones. We will then be able to compare the phase diagrams for the historical and parametric ES.

Phase diagram for parametric VaR and ES

I. Varga-Haszonits and I. Kondor: The instability of downside risk measures, *J. Stat. Mech.* P12007 doi: 10.1088/1742-5468/2008/12/P12007 (2008)



Adding linear constraints

In practice, portfolio optimization is always subject to some constraints on the allowed range of the weights, such as a ban on short selling and/or limits on various assets, industrial sectors, regions, etc. These constraints restrict the region over which the optimum is sought to a finite volume where no infinite fluctuations can appear. One might then think that under such constraints the instability discussed above disappears completely.

- This is not necessarily so. If we work in the vicinity of the phase boundary, sample to sample fluctuations in the weights will still be large, but the constraints will prevent the solution from running away to infinity. Instead, it will stick to the „walls" of the allowed region.
- For example, for a ban on short selling ($w_i > 0$) these walls will be the coordinate planes, and as N/T increases, more and more of the weights will become zero. This phenomenon is well known in portfolio optimization. (B. Scherer, R. D. Martin, Introduction to Modern Portfolio Optimization with NUOPT and S-PLUS, Springer, New York (2005))

- This spontaneous reduction of diversification occurs even for iid variables, where it is entirely due to estimation error and does not reflect any real structure of the objective function.
- In addition, for the next sample a completely different set of weights will become zero – the solution keeps jumping about on the walls of the allowed region.
- Clearly, in this situation the solution reflects the structure of the limit system (i.e. the portfolio manager’s beliefs), rather than the structure of the market. Therefore, whenever we are working in or close to the unstable region (which is almost always), the constraints only mask rather than cure the instability.

REGULARIZATION

A remedy from statistical learning theory: regularization

- Large fluctuations have to be penalized by adding a suitably chosen term to the objective function
- Choosing the regularizer: the L1 norm (the sum of the absolute values of the weights) is related to imposing a constraint on short selling. It prevents large longitudinal fluctuations, but leads to an increasing number of zero components as *N/T* increases.This may have advantages (reduces transaction costs), but is dangerous when the solution undergoes large sample fluctuations. (Brodie et al. applied L1 regularization to the variance as risk measure.)
- The advantages and dangers of sparse solutions.

Regularizing ES via the L2 norm

- This is related to support vector machines: S. Still and I. Kondor: Regularizing portfolio optimization, New Journal of Physics **12** 075034 (2010)
- The L2 norm (the sum of the squares of the weights) represent a diversification pressure and is logical to choose, given the tendency of shrinking diversification.
- Alternatively: adding the L2 term to Expected Shortfall can be interpreted as the effect of a linear market impact: F. Caccioli, S. Still, M. Marsili and I. Kondor: Optimal Liquidation Strategies Regularize Portfolio Selection, to appear in the European Journal of Finance, (2011)
- L2 does indeed take care of the instability: the phase transition disappears.

Market impact

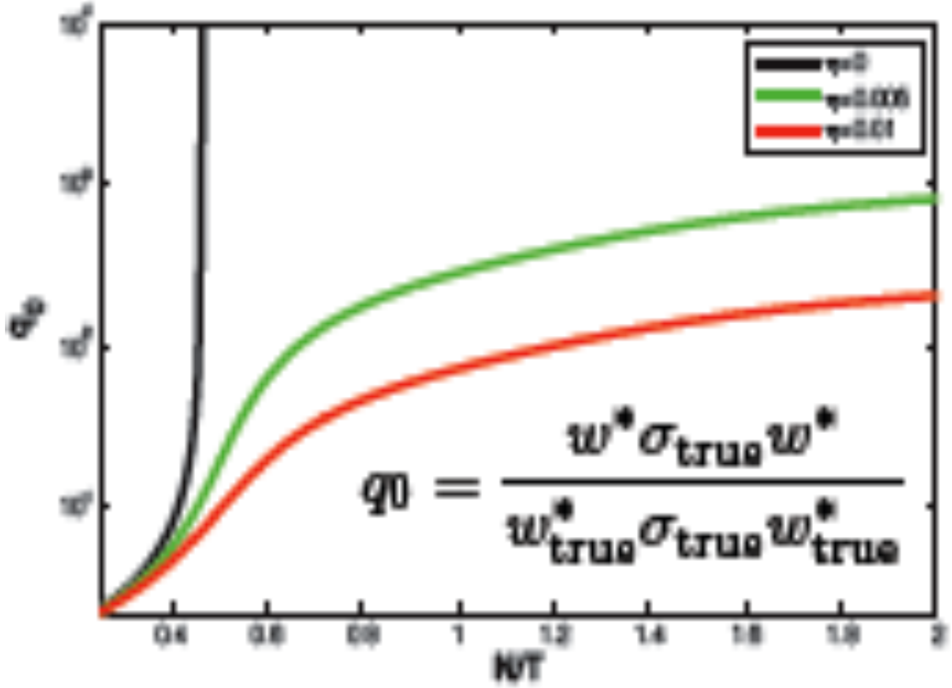
- Liquidation moves prices (specially in tail events)
- Realized prices in liquidating a position $\vec{w}$

$$\vec{p}_{\text{est}} = \vec{p}_{\text{now}} + \vec{x} - \vec{\psi}(\vec{w}) \qquad \text{e.g.} \quad \vec{\psi}(\vec{w}) = \eta \vec{w}$$

- Cash flow: $$\begin{aligned} c &= \vec{w} \cdot \vec{p}_{\text{est}} \\ &= \vec{w} \cdot \vec{p}_{\text{now}} + \vec{w} \cdot \vec{x} - \eta \vec{w} \cdot \vec{\psi}(\vec{w}) \end{aligned}$$

- Risk: $F(\vec{w} \cdot \vec{x} - \vec{w} \cdot \vec{\psi}) = F(\vec{w} \cdot \vec{x}) + \vec{w} \cdot \vec{\psi}(\vec{w})$
- Impact has same effect as regularizing

$$\text{e.g.} \quad \vec{\psi}(\vec{w}) = \eta \vec{w} |\vec{w}|^{p-2} \quad \Rightarrow \quad L_p \ \text{norm}$$



N/T
q_0
$$q0 = \frac{w^* \sigma_{\text{true}} w^*}{w^*_{\text{true}} \sigma_{\text{true}} w^*_{\text{true}}}$$
0.4
0.6
1
1.2
1.4
1.6
2

Closing remarks

Given the nature of the portfolio optimization task, one will typically work in that region of parameter space where sample fluctuations are large. Since the critical point where these fluctuations diverge depends on the risk measure, the confidence level, and on the method of estimation (historical or parametric), one must be aware of how close one's working point is to the critical boundary, otherwise one will be grossly misled by the unstable algorithm.

- Downside risk measures have been introduced, because they ignore positive fluctuations that investors are not supposed to be afraid of.

  Perhaps they should be: the downside risk measures display the instability described here which is basically due to a false arbitrage alert and may induce an investor to take very large positions on the basis of unreliable information stemming from finite samples.
- In a way, the recent crisis is a macroscopic example of such a folly.

- Regularization eliminates the instability and prevents the weight vector from undergoing large longitudinal fluctuations. The taming of transverse fluctuations is a task still to be attended to.
- The choice of the regularizer must be justified: here we pointed out that it represents a diversification pressure, but also argued that linear market impact also acts as an L2 regularizer.

THANK YOU!