THE INSTABILITY OF RISK MEASURES IS DUE TO THEIR UNBOUNDEDNESS

Imre Kondor

Parmenides Foundation, Pullach/Munich and London Mathematical Laboratory, London

Seminar talk at Sant'Anna, Pisa, 5 December, 2016.

Coworkers

Fabio Caccioli, UCL, London

Gábor Papp, Eötvös University, Budapest

Matteo Marsili, ICTP, Trieste

Susanna Still, University of Hawaii

István Varga-Haszonits, MSCI, Budapest

Motivation

- New market risk regulation (Basel Committee on Banking Supervision: Minimum capital requirements for market risk. Bank for International Settlements, 2016):

  - Expected Shortfall (ES) to replace Value at Risk

  - Confidence limit $\alpha = 0.975$

It is important to study the statistical properties of ES

- A wider machine learning context: ES optimization is closely related to ν-support vector regression.

Disclaimer

This talk is NOT meant to be a contribution to the ES vs. VaR debate.

The statistical problems we point out concerning ES are also shared by VaR.

Preliminary considerations

- Regulatory risk measures:
  - diagnostic tools
  - constraints on investment decisions
  - decision making tools

- Portfolio selection vs. risk measurement
- Out of sample prediction
- Portfolio selection is often a problem in high dimensional statistics.

- The role of $r = N/T$ (N is the number of different assets in the portfolio, T the length of available time series; dimension resp. sample size)
- At a critical value of r the estimation error diverges - phase transition.
- The fundamental cause of the instability is apparent arbitrage, due to the unbounded risk measure.
- We studied both historical and parametric estimates, but, for comparison, also the variance as risk measure.

Consider the simplest portfolio

$X = \sum_{i=1}^{N} w_i x_i$ with weights normalized as $\sum_{i=1}^{N} w_i = N.$

Prob. of loss $\ell(\{w_i\}, \{x_i\}) = -X$ to be smaller than a threshold ℓ_0

$$P(\{w_i\}, \ell_0) = \int \prod_{i=1}^{N} x_i dx_i p(\{x_i\}) \theta\left(\ell_0 - \ell(\{w_i\}, \{x_i\})\right)$$

Then the VaR:

$$\mathrm{VaR}_\alpha(\{w_i\}) = \min\{\ell_0 : P(\{w_i\}, \ell_0) \geq \alpha\}.$$

Expected Shortfall is the average loss above the VaR quantile:

$$\mathrm{ES}(\{w_i\}) = \frac{1}{1-\alpha} \int \Pi_i dx_i p(\{x_i\}) \ell(\{w_i\}, \{x_i\}) \theta(\ell(\{w_i\}, \{x_i\}) - \mathrm{VaR}_\alpha(\{w_i\}))$$

- The probability distribution of returns is not known, it must be inferred from observation of time series of returns.
- We generate these time series from i.i.d. normal distribution
- Optimizing ES can be reduced to the linear program:

minimize $$E(\epsilon, \{u_t\}) = (1-\alpha)T\epsilon + \sum_{t=1}^{T} u_t$$

under the constraints

$$u_t \geq 0 \ \ \forall\, t,$$

$$u_t + \epsilon + \sum_{i=1}^{N} x_{it} w_i \geq 0 \ \ \forall\, t,$$

$$\sum_i w_i = N.$$

- Note that the constraint on the expected return has been dropped – we are considering the global minimum.
- We consider the large N, large T limit, with their ratio $r = N/T$ fixed.
- This task can always be solved by numerical simulation, although for large N and T the task becomes computationally intensive, especially near the phase boundary, where fluctuations become very large.
- Alternatively: we have solved the linear program analytically, with tools borrowed from the statistical physics of disordered systems (replicas), and checked the results by simulations.

- For r = 0, we get the true value of ES (we call it $\mathrm{ES}^{(0)}$), and all the weights are 1.
- For r finite, the weights are Gaussian-distributed and fluctuate from sample to sample. The variance of their distribution is related to the out-of-sample estimate of ES as

$$\frac{\mathrm{ES}_{out}}{\mathrm{ES}^{(0)}} = \sqrt{\langle w^2 \rangle} = \sqrt{1+\sigma_w^2} = \sqrt{q_0}$$

- Susceptibility Δ, the sensitivity of the estimated weights to a small shift ξ in the returns:

$$\Delta = \left. \frac{\partial \langle w \rangle}{\partial \xi} \right|_{\xi=0}$$

- In-sample VaR: ϵ.

Replicas lead to the cost function:

$$
\begin{aligned}
F(\lambda, \epsilon, q_0, \Delta, \hat{q}_0, \hat{\Delta}) &= \lambda + \tau(1-\alpha)\epsilon - \Delta \hat{q}_0 - \hat{\Delta} q_0 \\
&+ \langle \min_w [V(w,z)] \rangle_z + \frac{\tau \Delta}{2\sqrt{\pi}} \int_{-\infty}^{\infty} ds e^{-s^2} g\left(\frac{\epsilon}{\Delta} + s\sqrt{\frac{2q_0}{\Delta^2}} \right)
\end{aligned}
$$

$$
V(w,z) = \hat{\Delta} w^2 - \lambda w - zw\sqrt{-2\hat{q}_0}
$$

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

Stationarity conditions

$$1 = \langle w^* \rangle_z$$

$$(1-\alpha) + \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} dse^{-s^2} g'\left(\frac{\epsilon}{\Delta} + s\sqrt{\frac{2q_0}{\Delta^2}}\right) = 0$$

$$\hat{\Delta} - \frac{1}{2r\sqrt{2\pi q_0}} \int_{-\infty}^{\infty} dse^{-s^2} sg'\left(\frac{\epsilon}{\Delta} + s\sqrt{\frac{2q_0}{\Delta^2}}\right) = 0$$

$$-\hat{q}_0 - 2\frac{\hat{\Delta} q_0}{\Delta} + \frac{1}{2r\sqrt{\pi}} \int_{-\infty}^{\infty} dse^{-s^2} g\left(\frac{\epsilon}{\Delta} + s\sqrt{\frac{2q_0}{\Delta^2}}\right) + \frac{(1-\alpha)}{r}\frac{\epsilon}{\Delta} = 0$$

$$\Delta = \frac{1}{\sqrt{-2\hat{q}_0}} \langle w^* z \rangle_z$$

$$q_0 = \left\langle w^{*2} \right\rangle_z .$$

Can be reduced:

$$r = \Phi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - \Phi\left(\frac{\epsilon}{\sqrt{q_0}}\right)$$

$$\alpha = \frac{\sqrt{q_0}}{\Delta}\left\{\Psi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - \Psi\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right\}$$

$$\frac{1}{2\Delta^2} + \frac{\alpha}{r}\frac{\epsilon}{\Delta} + \frac{1}{2}\frac{q_0}{\Delta^2} + \frac{1}{2r} = \frac{1}{r}\frac{q_0}{\Delta^2}\left\{W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right\}$$

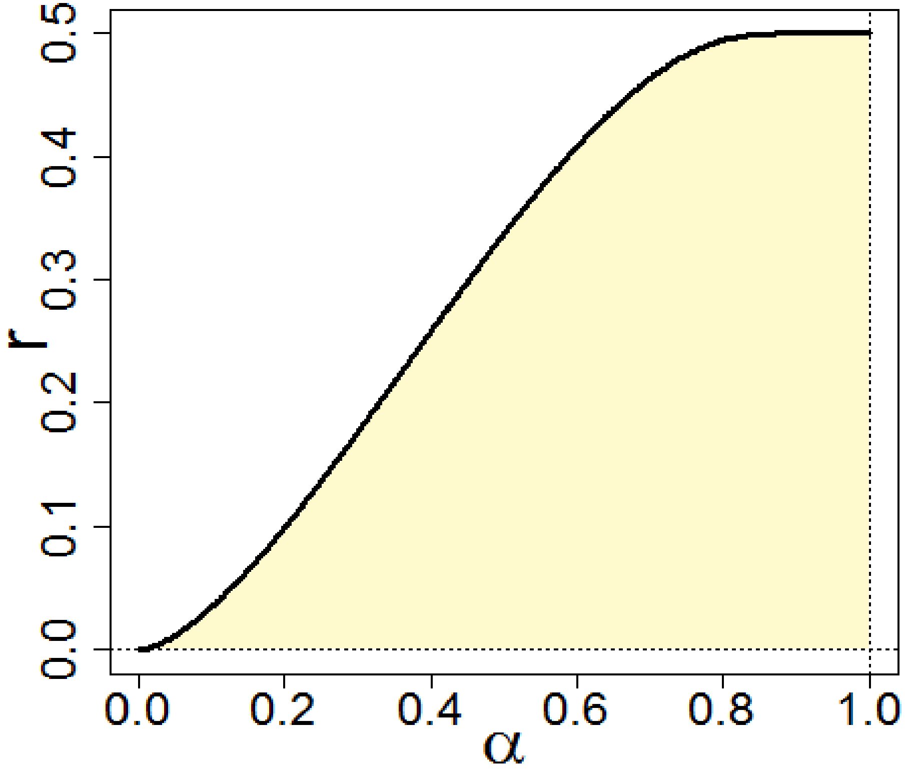
$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} dt\, e^{-t^2/2}$$

$$\Psi(x) = x\Phi(x) + \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

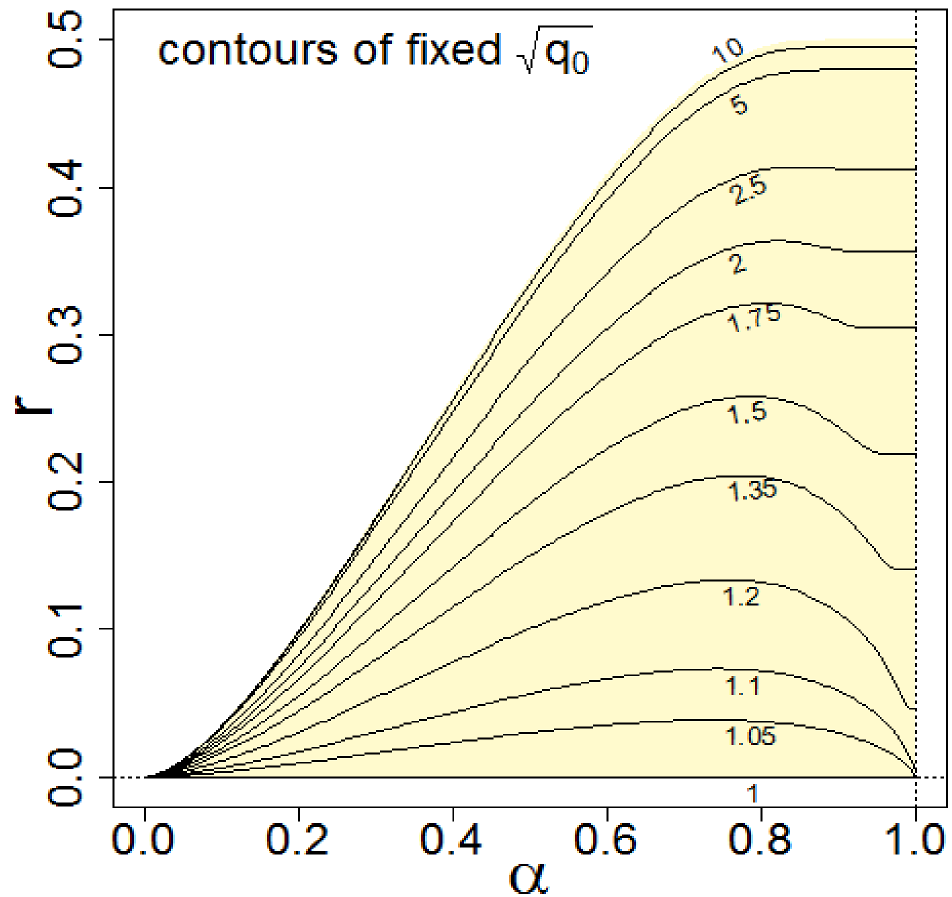
$$W(x) = \frac{x^2+1}{2}\Phi(x) + \frac{x}{2}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

The phase boundary:

The phase boundary of ES for i.i.d normal underlying returns. In the region below the phase boundary the optimization of ES is feasible and the estimation error is finite. Approaching the phase boundary from below, the estimation error diverges, and above the line optimization is no longer feasible.

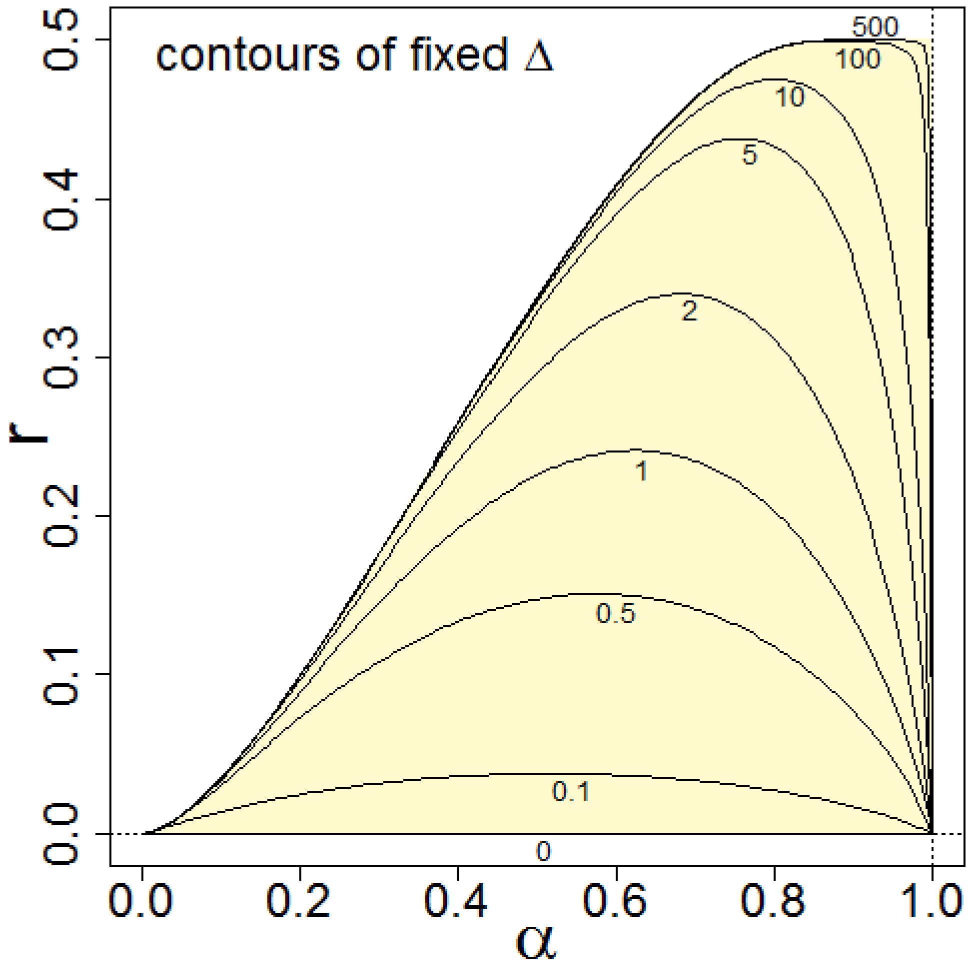


The contour map of the relative estimation error:



Contour lines of fixed $\sqrt{q_0}$. These curves are also the contour lines for the relative error for the out-of-sample estimate of ES.

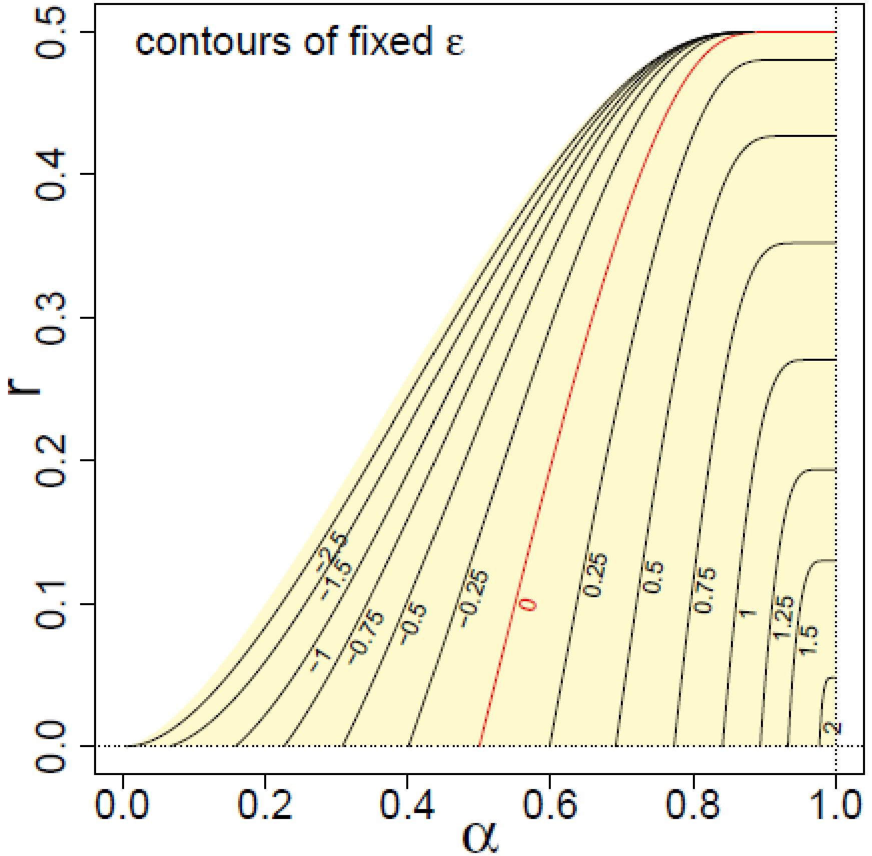
The contour map of the sensitivity to shifts in the returns:



Contour lines of fixed Δ, the susceptibility of the portfolio weights to small shifts in the returns. Δ is inversely proportional to the estimated in-sample ES.

Contour lines of fixed in-sample VaR

All three order parameters are singular at the phase boundary, and have an essential singularity at the upper right corner.
These maps allow one to determine the necessary amount of data to have a predetermined estimation error.



- Numerical results for historical ES: To have acceptably small estimation errors and susceptibilities one needs exceedingly large samples if *N* is not small.

estimation error ↓	α											
	0.7	0.8	0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.975	0.98
5%	26	27	33	35	37	39	43	47	53	64	72	83
10%	14	14	17	18	19	20	21	24	27	31	35	40
15%	10	10	12	12	13	13	14	16	18	20	22	25
20%	8	8	9	9	10	10	11	12	13	15	16	17
25%	6	6	7	8	8	8	9	9	10	11	12	12
50%	4	4	4	4	4	4	5	5	5	5	5	5

The table shows the (rounded) values of T/N that are needed to have a given estimation error for different values of the confidence level α. Even an estimation error of 25% requires samples 12 times larger than the number of items in the portfolios at the confidence level α=0.975 proposed by regulation.

- According to numerical experiments, fat tailed distributions need even more data than those in the table.

- Parametric estimates require less data than the historical estimates.

estimation error ↓	α											
	0.7	0.8	0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.975	0.98
5%	19	16	14	14	14	14	13	13	13	13	13	13
10%	10	9	8	8	7	7	7	7	7	7	7	7
15%	7	6	5	5	5	5	5	5	5	5	5	5
20%	6	5	4	4	4	4	4	4	4	4	4	4
25%	5	4	4	4	4	4	3	3	3	3	3	3
50%	3	3	2	2	2	2	2	2	2	2	2	2

This table shows the (rounded) values of T/N needed to have a given estimation error for different values of the confidence level α in the case of parametric estimation.

- For α close to 1 the data requirement of parametric VaR is essentially the same as that of the parametric ES.
- If variance is the risk measure the required data is slightly less than the $\alpha = 0.975$ parametric ES values (at 5% e.g. $T/N = 11$).

Regularization: adding a term to the cost function with the purpose to rein in large fluctuations at the price of introducing bias. The ℓ_2-regularizer represents a diversification pressure. Alternatively: regularization takes into account market impact.

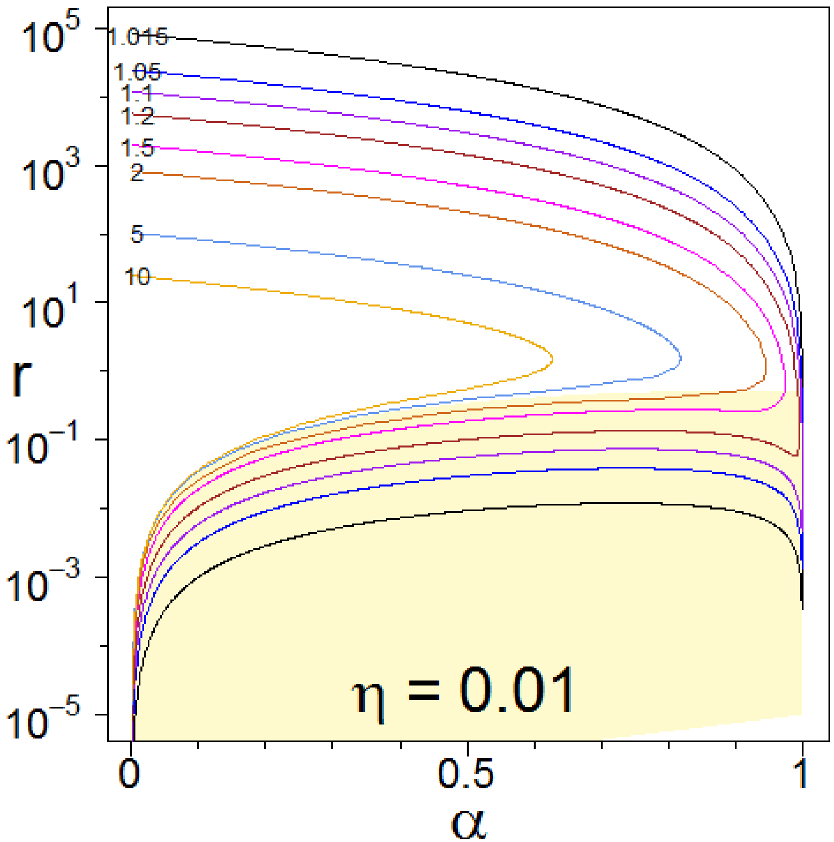
The modified optimization task is:

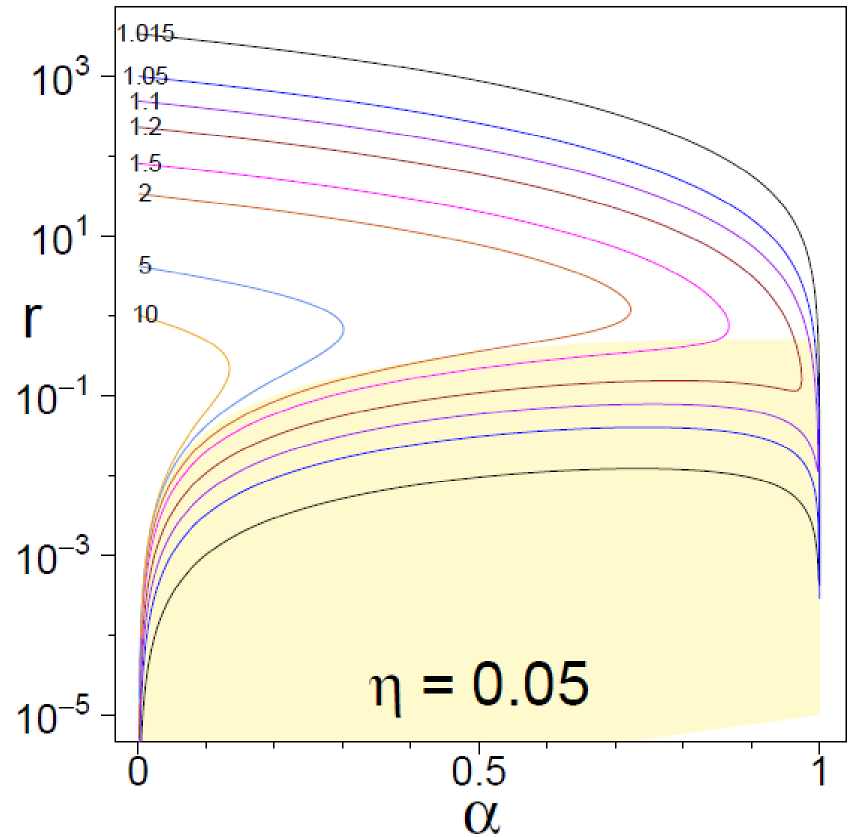
$$\min_{\vec{w},\vec{u},\epsilon}\left[(1-\alpha)T\epsilon+\sum_{t=1}^{T}u_t+\eta\sum_i w_i^2\right],$$

$$\text{s.t.}\qquad \vec{w}\cdot\vec{x}_t+\epsilon+u_t\geq 0;\quad u_t\geq 0;\quad \forall t,$$

$$\sum_i w_i = N,$$

Relative estimation error for $\boldsymbol{\ell}_2$-regularized ES

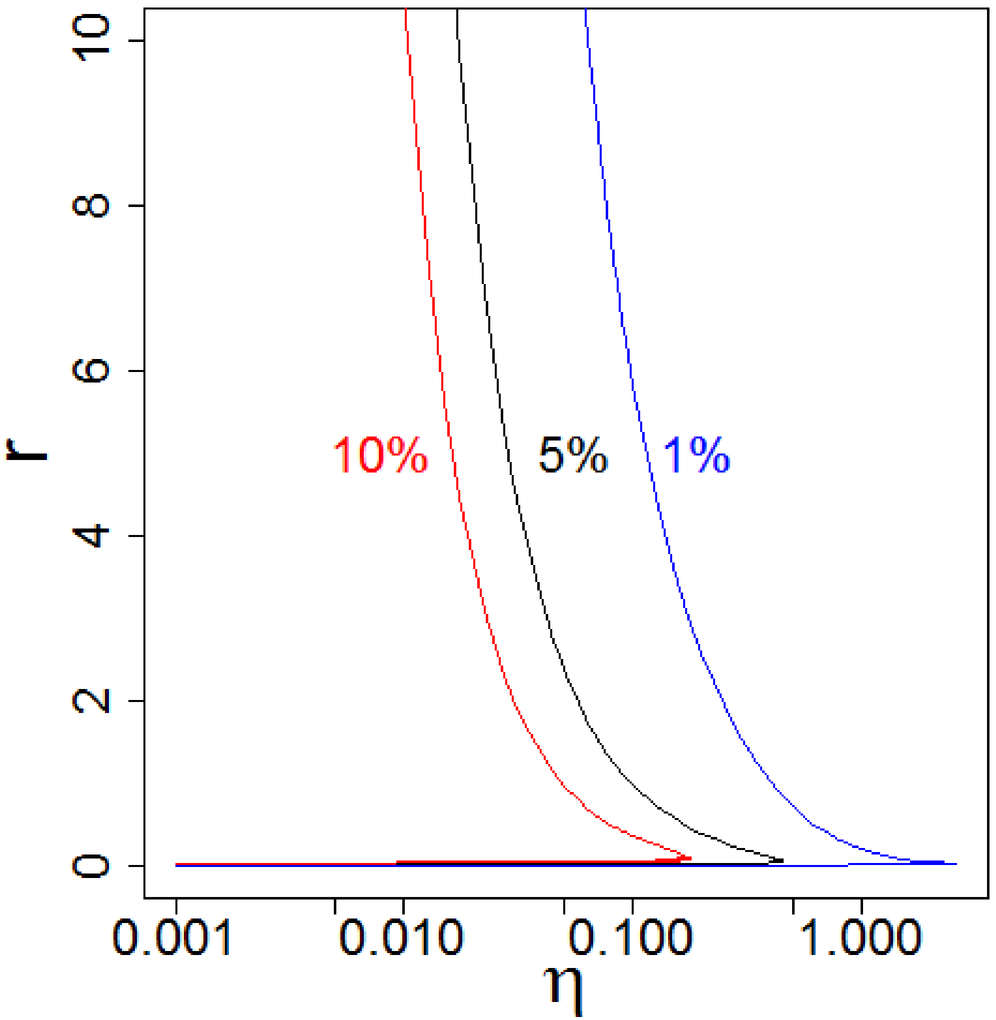




Contour plot for fixed $\sqrt{q_0}$ on the $\alpha -\ r$ plane at η=0.01 (left) and η=0.05 (right). Note the log scale on the vertical axis. (It would be hard to reproduce such sharp features by simulation.)

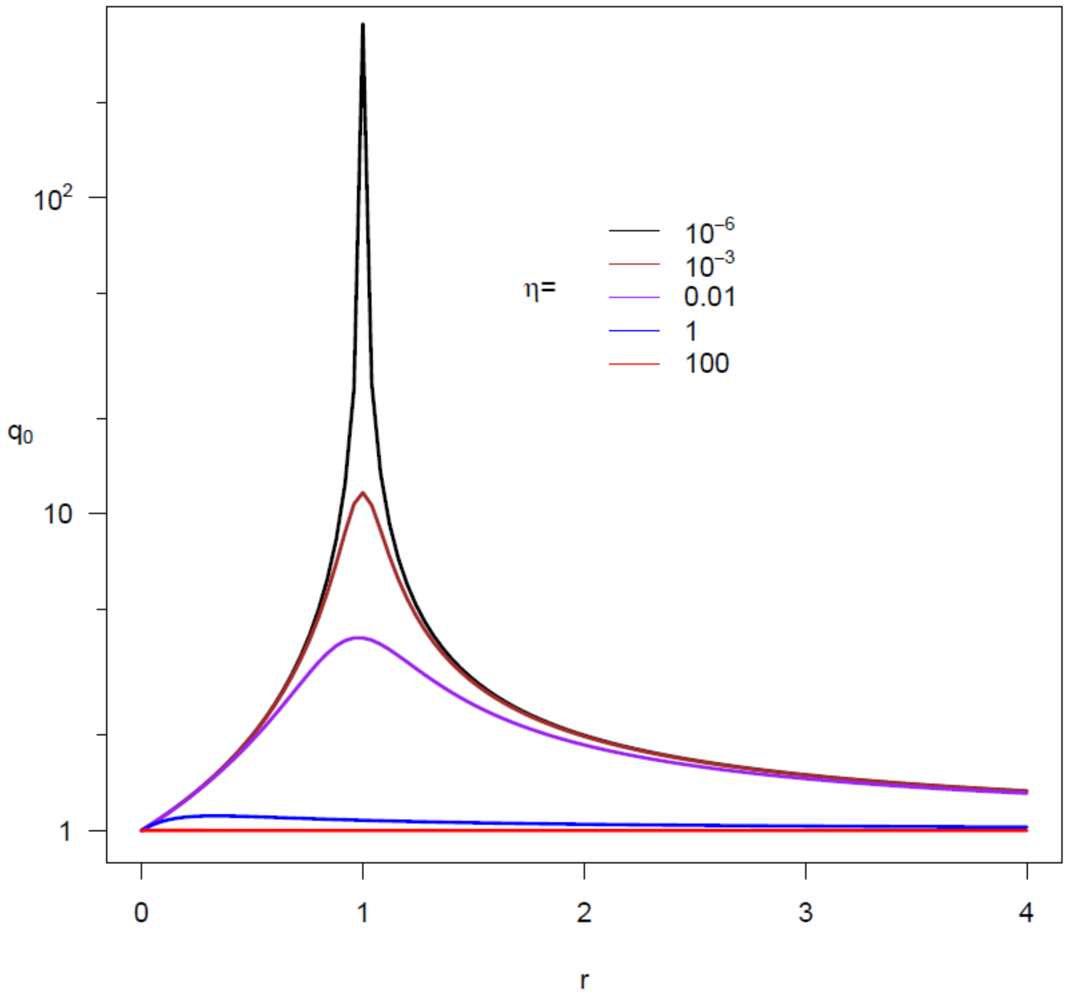
Variance-bias- trade-off for ℓ_2-regularized ES

The value of r=N/T vs. the coefficient of the regularizer for a fixed value of the confidence limit α=0.975 and three different values (1%, 5% and 10%) of the relative estimation error. The data-dominated and bias-dominated regions can be clearly distinguished in these curves: in the range of small r's the curves depend on the strength of the regularizer very weakly, while for larger r's (but still below the phase boundary) the regularizer starts to dominate. Data are crowded out by bias in this region. Note the sharp transition between these two regimes.

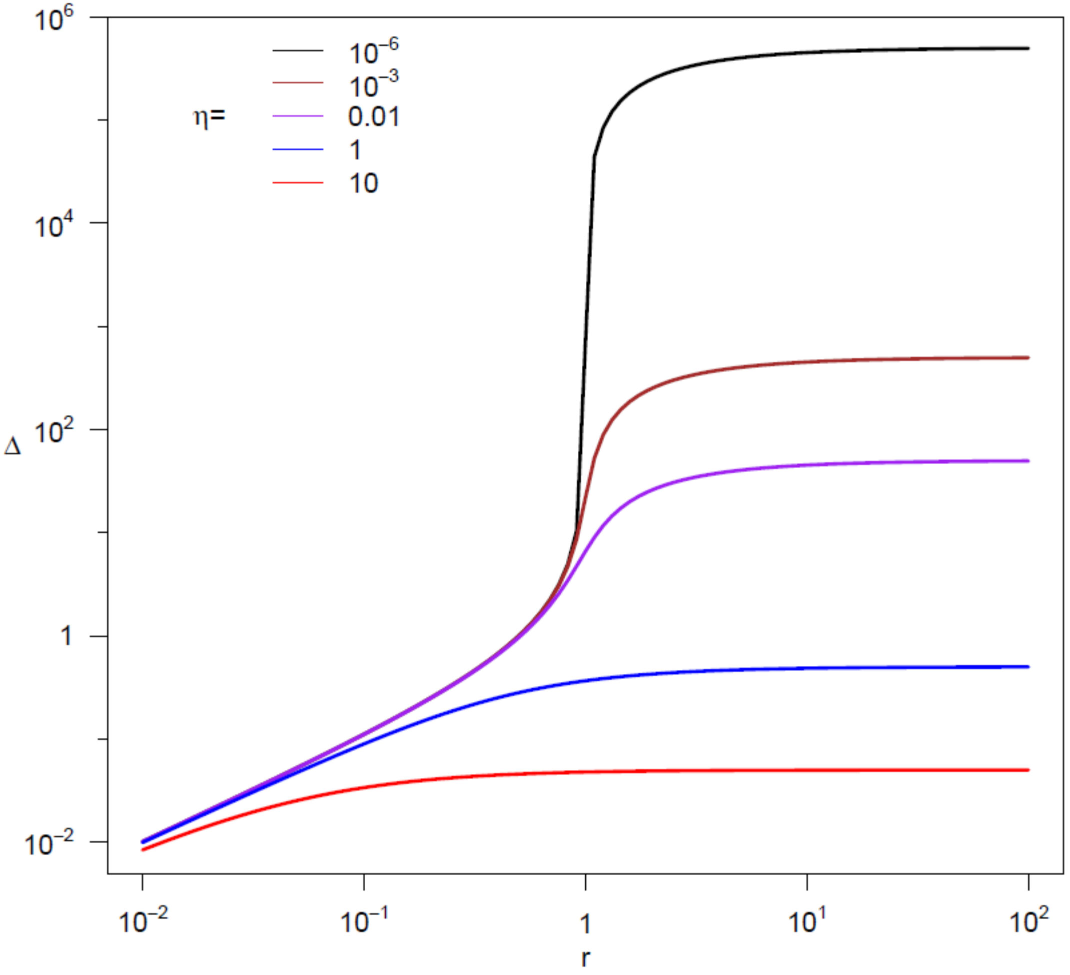


Comparison with the variance as risk measure

Relative error

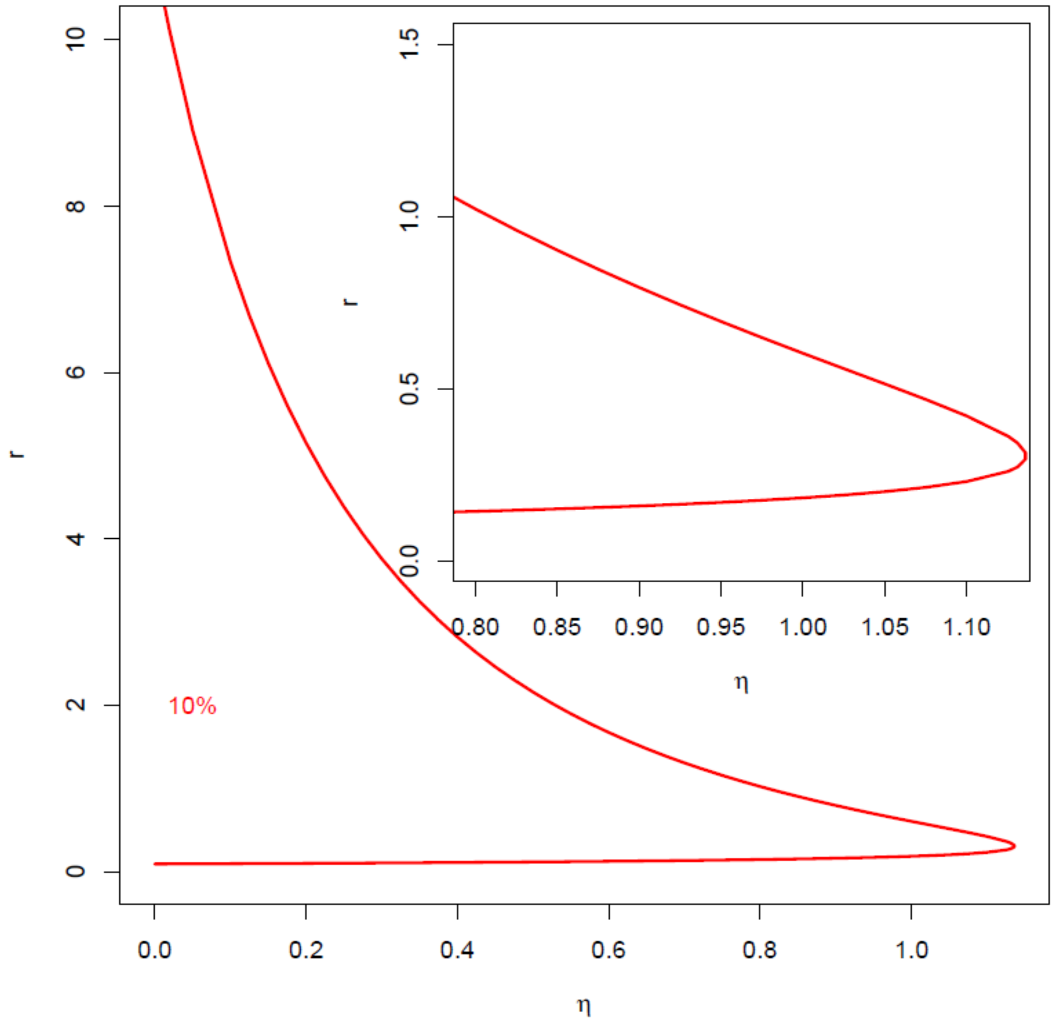


Susceptibility

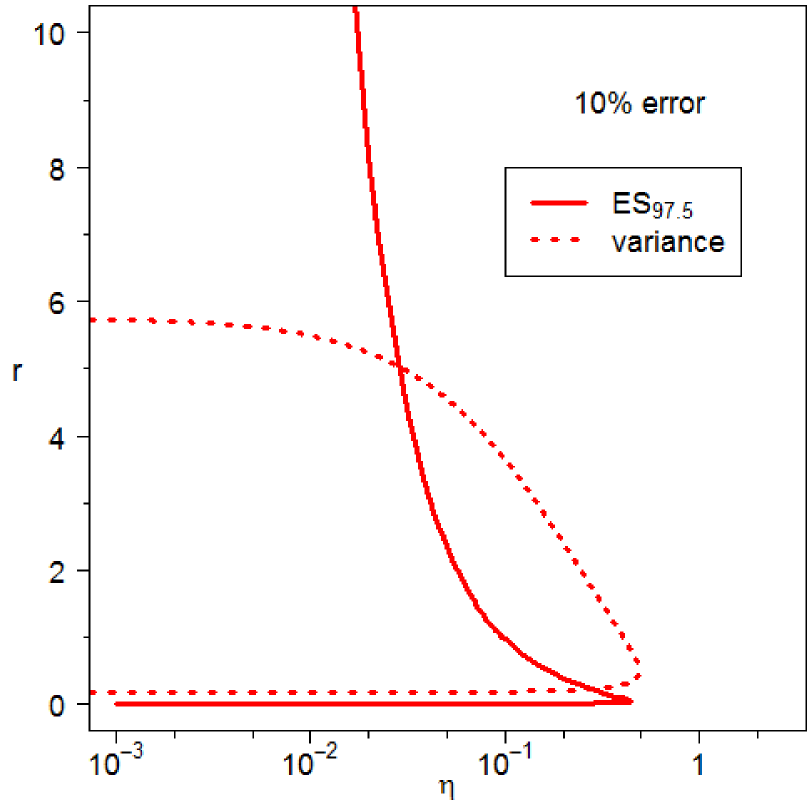


Variance-bias trade-off

Note how narrow is the transition region again!



Variance-bias trade-off: The transition zone is wider for the variance, but in terms of *N/T* smaller than a factor four even there.



Unconstrained variance

- Critical point at *r=N/T=1*, where the covariance matrix starts to have zero eigenvalues.
- Estimation error and susceptibility diverge, cost function tends to zero.
- For *N>T*, *N-T* zero modes appear, cost function gets fixed at zero, continuum of solutions.
- Large compensating positions – high leverage.
- Ready-made solvers keep finding a definite solution along the diagonal due to built-in ℓ_2 regularizer.
- Dangerous to use ready-made tools.

Variance with no short positions

- Ban on short selling is a frequent constraint
- It is a special ℓ_1 regularizer, leads to sparse solution
- Solution exist up to r=2.
- wCw is sum of T terms. If the number of zero modes N-T > T, we can use the zero modes to make every term zero. So the cost function becomes flat above r=2 now.

Phase transition at $r = 2$

- Cost function goes to zero, susceptibility diverges, but estimation error remains finite – no obvious danger signal.
- The no-short condition reins in longitudinal fluctuations, but large transverse fluctuations remain.
- Standard solvers (in R, Matlab) keep finding stable solution even where there is none.
- Banks often use black boxes for asset and risk management – a component of systemic risk.

CONCLUSION

- Insufficient data cannot support good decisions.
- „Sufficient” means irrealistically large data for ES.
- The usual remedy is regularization. We find two clearly distinguishable regions where the data resp. the bias dominate. The transition zone between them, where a meaningful trade-off can take place, is very narrow.
- Regularization works fine in natural sciences or machine learning applications. Why not in finance? Unbounded risk measures invite high leverage, the mother of all fianancial risks.

Some references

- I. Kondor, Sz. Pafka, G. Nagy: Noise sensitivity of portfolio selection under various risk measures, *Journal of Banking and Finance*, **31**, 1545-1573 (2007).
- S. Ciliberti, I. Kondor, M. Mezard: On the Feasibility of Portfolio Optimization under Expected Shortfall, *Quantitative Finance*, **7**, 389-396 (2007)
- 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)
- S. Still and I. Kondor: Regularizing portfolio optimization, New Journal of Physics, **12,** 075034 Doi: 10.1088/1367-2630/12/7/075034 (2010).
- F. Caccioli, S. Still, M. Marsili and I. Kondor: Optimal Liquidation Strategies Regularize Portfolio Selection, European Journal of Finance, **19**, 554-571 (2013), Doi: 10.1080/1351847X.2011.601661, http://arXiv:1004.4169v1 [q-fin.PM] (2010).
- F. Caccioli, I. Kondor, M. Marsili and S. Still: L_p regularized portfolio optimization, working paper, arXiv:1404.4040[q-fin.PM], http://ssrn.com/abstract=2425326 (2015)
- F. Caccioli, I. Kondor and G. Papp: Portfolio optimization under Expected Shortfall: Contour maps of estimation error, http://arxiv.org/abs/1510.04943 (2015)
- F. Caccioli, G. Papp and I. Kondor: Fluctuation-bias trade-off in portfolio optimization under Expected Shortfall with l_2 regularization, http://arXiv:1602.08297v1 [q-fin.PM] (2016)

THANK YOU!