Journal of Statistical Mechanics:
Theory and Experiment

PAPER: INTERDISCIPLINARY STATISTICAL MECHANICS • OPEN ACCESS

Bias-variance trade-off in portfolio optimization under expected shortfall with ℓ_2 regularization

To cite this article: Gábor Papp *et al J. Stat. Mech.* (2019) 013402

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 128.41.35.142 on 06/02/2019 at 17:23

OPEN ACCESS

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

PAPER: Interdisciplinary statistical mechanics

Bias-variance trade-off in portfolio optimization under expected shortfall with ℓ_2 regularization

Gábor Papp[1], **Fabio Caccioli**[2,3,5] **and Imre Kondor**[4,5,6]

[1] Eötvös Loránd University, Institute for Physics, Budapest, Hungary
[2] Department of Computer Science, University College London, London, WC1E 6BT, United Kingdom
[3] Systemic Risk Centre, London School of Economics and Political Sciences, London, United Kingdom
[4] Parmenides Foundation, Pullach, Germany
[5] London Mathematical Laboratory, London, United Kingdom
[6] Complexity Science Hub, Vienna, Austria
E-mail: f.caccioli@ucl.ac.uk

Received 5 July 2018
Accepted for publication 11 November 2018
Published 4 January 2019

Online at stacks.iop.org/JSTAT/2019/013402
https://doi.org/10.1088/1742-5468/aaf108

Abstract. The optimization of a large random portfolio under the expected shortfall risk measure with an ℓ_2 regularizer is carried out by analytical calculation for the case of uncorrelated Gaussian returns. The regularizer reins in the large sample fluctuations and the concomitant divergent estimation error, and eliminates the phase transition where this error would otherwise blow up. In the data-dominated region, where the number N of different assets in the portfolio is much less than the length T of the available time series, the regularizer plays a negligible role even if its strength η is large, while in the opposite limit, where the size of samples is comparable to, or even smaller than the number of assets, the optimum is almost entirely determined by the regularizer. We construct the contour map of estimation error on the N/T versus η plane and find that for a given value of the estimation error the gain in N/T due to the regularizer can reach a factor of about four for a sufficiently strong regularizer.

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: quantitative finance, risk measure and management, cavity and replica method

Contents

1. Introduction 2
2. The optimization of regularized ES 3
3. Results 7
4. Conclusion 12
Acknowledgments 13
References 13

1. Introduction

The current international market risk regulation measures risk in terms of expected shortfall (ES) [1]. In order to decrease their capital requirements, financial institutions have to optimize the ES of their trading book.

Optimizing a large portfolio may be difficult under any risk measure, but becomes particularly hard in the case of Value at Risk (VaR) and ES that discard a large part of the data except those at or above a high quantile. Without some kind of regularization, this leads to a phase transition at a critical value r_c of the ratio $r = N/T$ where N is the dimension of the portfolio (the number of different assets or risk factors) and T is the sample size (the length of the available time series). This critical ratio depends on the confidence level α that determines the VaR threshold above which the losses are to be averaged to obtain the ES. Beyond r_c it is impossible to carry out the optimization, and upon approaching this critical value from below the estimation error increases without bound.

The estimation error problem of portfolio selection has been the subject of a large number of works, [2–12] are but a small selection from this vast literature. The critical behavior and the locus of the phase boundary separating the region where the optimization is feasible from the one where it is not has also been studied in a series of papers [13–17].

In the present note we discuss the effect of adding an ℓ_2 regularizer to the ES risk measure. As noted in [18] and [19], the optimization problem so obtained is equivalent to one of the variants of support vector regression (ν-SVR) [20], therefore its study is of interest also for machine learning, independently of the portfolio optimization context. Concerning its specific financial application, ℓ_2 regularization may have two different interpretations. First, ℓ_2 has the tendency to pull the solution towards the diagonal, where all the weights are the same. As such, ℓ_2 represents a diversification pressure [21, 22] that may be useful in a situation where, e.g. the market is dominated by a small number of blue chips. Alternatively, the portfolio manager may wish to take

J. Stat. Mech. (2019) 013402

into account the market impact of the future liquidation of the portfolio already at its construction. As shown in [17], market impact considerations lead one to regularized portfolio optimization, with ℓ_2 corresponding to linear impact.

In this paper we carry out the optimization of ℓ_2-regularized ES analytically in the special case of i.i.d. Gaussian distributed returns, in the limit where both the dimension and the sample size are large, but their ratio $r = N/T$ is fixed. The calculation will be performed by the method of replicas borrowed from the statistical physics of disordered systems [23]. The present work extends a previous paper [24] by incorporating the regularizer. By preventing the phase transition from taking place, the regularizer fundamentally alters the overall picture (in this respect, the role of the regularizer is analogous to that of an external field coupled to the order parameter in a phase transition). As the technical details of the replica calculation have been laid out in [13] and, in a somewhat different form, in [17], we do not repeat them here. Instead we just recall the setup of the problem and focus on the most important result: the relative estimation error (closely related to the out-of-sample estimator of ES) as a function of $r = N/T$ and the strength of the regularizer η.

Our results exhibit a clear distinction between the region in the space of parameters where data dominate and the regularizer plays a minor role, from the one where the data are insufficient and the regularizer stabilizes the estimate at the price of essentially suppressing the data. Thereby, our results provide a clean and explicit example of what statisticians call the bias-variance trade-off that lies at the heart of the regularization procedure. We find that the transition between the data-dominated regime and the bias-dominated one is rather sharp, and it is only around this transition that an actual trade-off takes place. Following the curves of fixed estimation error on the $r - \eta$ plane we can see that r increases with increasing η by a factor of approximately 4. Beyond this point the contour lines turn back and we go over onto a branch of the contour line where the optimization is determined by the regularizer rather than the data. The plan of the rest of the paper is as follows. In section 2 we present the formalism, in section 3 display our results and in section 4 draw our conclusions.

2. The optimization of regularized ES

The simple portfolios we consider here are linear combinations of N risk factors, with returns x_i and weights w_i, $i = 1, 2, ..., N$:

$$X = \sum_{i=1}^{N} w_i x_i. \tag{1}$$

The weights will be normalized such that their sum is N, instead of the customary 1:

$$\sum_{i=1}^{N} w_i = N. \tag{2}$$

The motivation for choosing this normalization is that we wish to have weights of order unity, rather than $1/N$, in the limit $N \to \infty$. Apart from the budget constraint,

J. Stat. Mech. (2019) 013402

the weights will not be subject to any other condition. In particular, they can take any real value, that is we are allowing unlimited short positions. We do not impose the usual constraint on the expected return on the portfolio either, so we are looking for the global minimum risk portfolio. This setup is motivated by simplicity, but we note that tracking a benchmark requires exactly this kind of optimization to be performed.

The probability for the loss $\ell(\{w_i\},\{x_i\}) = -X$ to be smaller than a threshold ℓ_0 is:

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

where $p(\{x_i\})$ is the probability density of the returns, and $\theta(x)$ is the Heaviside function: $\theta(x) = 1$ for $x > 0$, and zero otherwise. The VaR at confidence level α is then defined as:

$$\mathrm{VaR}_\alpha(\{w_i\}) = \min\{\ell_0 : P(\{w_i\},\ell_0) \geqslant \alpha\}. \tag{3}$$

Expected Shortfall is the average loss beyond the VaR quantile:

$$\mathrm{ES}(\{w_i\}) = \frac{1}{1-\alpha} \int \Pi_i \mathrm{d}x_i p(\{x_i\}) \ell(\{w_i\},\{x_i\}) \theta(\ell(\{w_i\},\{x_i\}) - \mathrm{VaR}_\alpha(\{w_i\})). \tag{4}$$

Portfolio optimization seeks to find the optimal weights that make the above ES minimal subject to the budget constraint (2). Instead of dealing directly with ES, Rockafellar and Uryasev [25] proposed to minimize the related function

$$F_\alpha(\{w_i\},\epsilon) = \epsilon + \frac{1}{1-\alpha} \int \Pi_i \mathrm{d}x_i p(\{x_i\}) \left[\ell(\{w_i\},\{x_i\}) - \epsilon\right]^+ \tag{5}$$

over the variable ϵ and the weights w_i:

$$\mathrm{ES}(\{w_i\}) = \min_\epsilon F_\alpha(\{w_i\},\epsilon), \tag{6}$$

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

The probability distribution of the returns is not known, so one can only sample it, and replace the integral in (4) by time-averaging over the discrete observations. Rockafellar and Uryasev [25] showed that the optimization of the resulting loss function can be reduced to the following linear programming task: minimize the cost function

$$E(\epsilon,\{u_t\}) = (1-\alpha)T\epsilon + \sum_{t=1}^{T} u_t \tag{7}$$

under the constraints

$$\begin{aligned} &u_t \geqslant 0 \ \forall\ t, \\ &u_t + \epsilon + \sum_{i=1}^{N} x_{it} w_i \geqslant 0 \ \forall\ t, \\ &\text{and} \quad \sum_i w_i = N. \end{aligned} \tag{8}$$

It is convenient to introduce the regularizer at this stage, by adding the ℓ_2-norm to the loss function [17]:

J. Stat. Mech. (2019) 013402

$$\min_{\vec{w},\vec{u},\epsilon}\left[(1-\alpha)T\epsilon+\sum_{t=1}^{T}u_t+\eta\sum_i w_i^2\right], \tag{9}$$

$$\text{s.t.}\quad \vec{w}\cdot\vec{x}_t+\epsilon+u_t\geqslant 0;\quad u_t\geqslant 0;\quad \forall t, \tag{10}$$

$$\sum_i w_i=N, \tag{11}$$

where ϵ and $\vec{u}$ are auxiliary variables, and the coefficient η sets the strength of the regularization.

As the constraint on the expected return has been omitted, we are seeking the global optimum of the portfolio here. If the returns x_{it} are i.i.d. Gaussian variables and $N,T\to\infty$ with $r=N/T$ fixed, the method of replicas allows one to reduce the above optimization task in $N+T+1$ variables to the optimization of a cost function depending on only six variables, the so-called order parameters [13, 17]:

$$F(\lambda,\epsilon,q_0,\Delta,\hat{q}_0,\hat{\Delta})=\lambda+\frac{1}{r}(1-\alpha)\epsilon-\Delta\hat{q}_0-\hat{\Delta}q_0 + \langle\min_w\left[V(w,z)\right]\rangle_z+\frac{\Delta}{2r\sqrt{\pi}}\int_{-\infty}^{\infty}\mathrm{d}s\mathrm{e}^{-s^2}g\left(\frac{\epsilon}{\Delta}+s\sqrt{\frac{2q_0}{\Delta^2}}\right), \tag{12}$$

where

$$V(w,z)=\hat{\Delta}w^2-\lambda w-zw\sqrt{-2\hat{q}_0}+\eta w^2, \tag{13}$$

$\langle\cdot\rangle_z$ is an average over the standard normal variable z, and

$$g(x)=\begin{cases}0, & x\geqslant 0\\ x^2, & -1\leqslant x\leqslant 0\\ -2x-1, & x<-1\end{cases}. \tag{14}$$

One can readily see that the stationarity conditions are:

$$1=\langle w^*\rangle_z \tag{15}$$

$$(1-\alpha)+\frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty}\mathrm{d}s\ \mathrm{e}^{-s^2}g'\left(\frac{\epsilon}{\Delta}+s\sqrt{\frac{2q_0}{\Delta^2}}\right)=0 \tag{16}$$

$$\hat{\Delta}-\frac{1}{2r\sqrt{2\pi q_0}}\int_{-\infty}^{\infty}\mathrm{d}s\ \mathrm{e}^{-s^2}sg'\left(\frac{\epsilon}{\Delta}+s\sqrt{\frac{2q_0}{\Delta^2}}\right)=0 \tag{17}$$

$$-\hat{q}_0-2\frac{\hat{\Delta}q_0}{\Delta}+\frac{1}{2r\sqrt{\pi}}\int_{-\infty}^{\infty}\mathrm{d}s\ \mathrm{e}^{-s^2}g\left(\frac{\epsilon}{\Delta}+s\sqrt{\frac{2q_0}{\Delta^2}}\right)+\frac{(1-\alpha)}{r}\frac{\epsilon}{\Delta}=0 \tag{18}$$

$$\Delta=\frac{1}{\sqrt{-2\hat{q}_0}}\langle w^*z\rangle_z \tag{19}$$

$$q_0=\left\langle w^{*2}\right\rangle_z, \tag{20}$$

J. Stat. Mech. (2019) 013402

where the variable w^* is that value of the weight that minimizes the 'potential' V in (13). (The prime means derivative with respect to the argument.)

Three of the order parameters are easily eliminated and the integrals can be reduced to the error function and its integrals by repeated integration by parts, as in [24]. Finally, one ends up with three equations to be solved:

$$r\,(1-2\eta\Delta) = \Phi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - \Phi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) \tag{21}$$

$$\alpha = \frac{\sqrt{q_0}}{\Delta}\left\{\Psi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - \Psi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right)\right\} \tag{22}$$

$$\frac{1}{2\Delta^2} + \frac{\alpha}{r}\frac{\epsilon}{\Delta} + \frac{q_0}{2\Delta^2} + \frac{1}{2r} - \frac{2\eta q_0}{\Delta^2} = \frac{q_0}{r\Delta^2}\left\{W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right)\right\}, \tag{23}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{d}t \mathrm{e}^{-t^2/2}, \tag{24}$$

$$\Psi(x) = x\,\Phi(x) + \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}, \tag{25}$$

$$W(x) = \frac{x^2+1}{2}\Phi(x) + \frac{x}{2}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}. \tag{26}$$

These functions are closely related to each other: $\Phi(x)$ is the derivative of $\Psi(x)$ and $\Psi(x)$ is the derivative of $W(x)$.

As explained in [24], each of the three remaining order parameters in the above set of equations, q_0, Δ, and ϵ has a direct financial meaning: Δ is related to the in-sample estimator of ES (and also to the second derivative of the cost function F with respect to the Lagrange multiplyer λ associated with the budget constraint) and ϵ is the in-sample VaR of the portfolio optimized under the ES risk measure. Our present concern is the order parameter q_0, which is a measure of the out-of-sample estimator of ES. As shown in [24], if $\mathrm{ES}_{\mathrm{out}}$ is the out-of-sample estimate of ES based on samples of size T, and $\mathrm{ES}^{(0)}$ is its true value (that would obtain for N finite and $T \to \infty$), then

$$\frac{\mathrm{ES}_{\mathrm{out}}}{\mathrm{ES}^{(0)}} = \sqrt{q_0}, \tag{27}$$

that is $\sqrt{q_0} - 1$ is the relative estimation error of the out-of-sample estimate.

The task now is to solve the stationary conditions and get the cost function by substituting the solutions back into equation (12). The in-sample value of Expected Shortfall is simply related to the cost function as:

$$\mathrm{ES} = \frac{rF}{1-\alpha}. \tag{28}$$

J. Stat. Mech. (2019) 013402

The fundamental cause of the divergence of estimation error in the original, non-regularized problem is that ES as a risk measure is not bounded from below. In finite samples it can happen that one of the assets, or a combination of assets, dominates the others (i.e. produces a larger return than the others in the given sample), thereby leading to an apparent arbitrage: one can achieve an arbitrarily large gain (an arbitrarily large negative ES) by going very long in the dominant asset and correspondingly short in the others [13, 26]. It is evident that this apparent arbitrage is a mere statistical fluctuation, but along a special curve in the $r - \alpha$ plane this divergence occurs with probability one [13]. As a result, the estimation error will diverge along the phase boundary shown in figure 1. Note that in high-dimensional statistics where regularization is routinely applied the loss is always bounded both from above and below. The setting in the present paper is, therefore, very different from the customary setup, which explaines the unusual results.

The purpose of regularization is to penalize the large fluctuations of the weight vector, thereby eliminating this phase transition.

Since ES is a piecewise linear function of the weights, the quadratic regularizer will overcome excessive fluctuations, no matter how small the coefficient η is. Deep inside the region of stability (shown by pale yellow in figure 1), a weak regularizer (small η) will modify the behavior of various quantities very little. In contrast, close to the phase boundary, and especially in the vicinity of the point $\alpha = 1$, $r = 0.5$, where the solution has an essential singularity, the effect of even a small η is very strong, and beyond the yellow region, where originally there was no solution, the regularizer will dominate the scene. In the region where the solution is stable even without the regularizer, $r = N/T$ is small, which means we have an abundance of data. We call this region the data-dominated region. In the presence of the regularizer we will find finite solutions also far beyond the phase boundary, but here the regularizer dominates the data, so we can call this domain the bias-dominated region.

3. Results

The solution of the stationarity conditions can be obtained with the help of a computer. In the following, we present the numerical solutions for the relative estimation error. The results will be displayed by constructing the contour map of this quantity, which will allow us to make a direct comparison between our present results and those in [24].

In figure 2 we recall the contour map of the relative estimation error of ES without regularization.

As can be seen, without regularization the constant q_0 curves are all inside the feasible region. For larger and larger values of q_0 the corresponding curves run closer and closer to the phase boundary, along which the estimation error diverges. Note that the phase boundary becomes flat, with all its derivatives vanishing, at the upper right corner of the feasible region: there is an essential singularity at the point $\alpha = 1, r = 0.5$.

The estimation error problem is very clearly illustrated in this figure: the curves corresponding to an acceptably small relative error are the lowest ones among the q_0 contour lines, and the value of $r = N/T$ corresponding to a confidence limit α in the

J. Stat. Mech. (2019) 013402

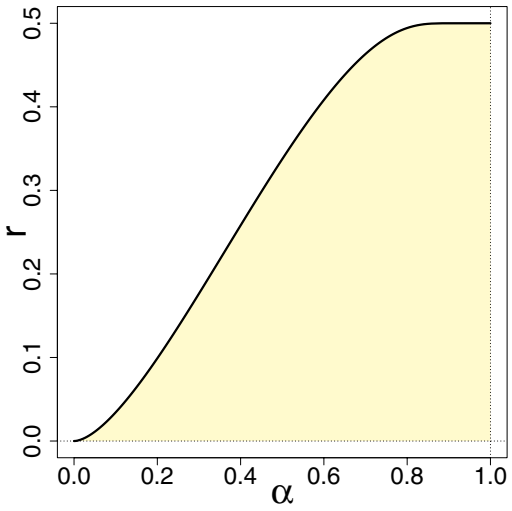

Figure 1. The phase boundary of unregularized ES for i.i.d. normal underlying returns. In the region below the phase boundary the optimization of ES is feasible and the estimation error is finite. Approaching the phase boundary from below, the estimation error diverges, and above the line optimization is no longer feasible.

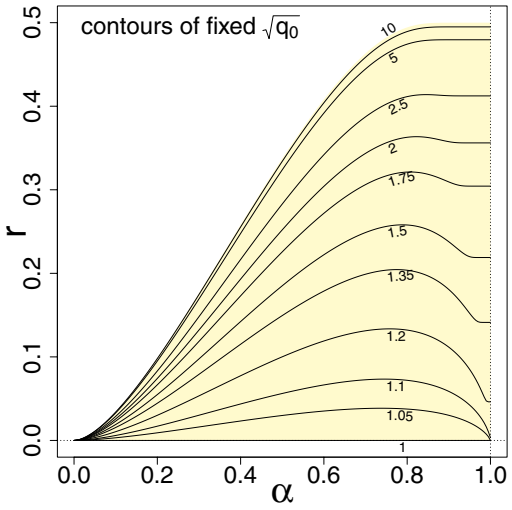

Figure 2. Contour lines of fixed $\sqrt{q_0}$ in the absence of regularization. These curves are also the contour lines for the relative error for the out-of-sample estimate of ES.

vicinity of 1 (such as the regulatory value $\alpha = 0.975$) are extremely small on these low lying curves. These small values of r require an unrealistically large sample size T if N is not small. For example, at the regulatory value of $\alpha = 0.975$, to achieve an estimation error smaller than 5%, for a portfolio with $N = 100$ assets one would need a time series of more than 7200 data points [24].

Let us see how regularization reorganizes the set of constant q_0 curves. Figures 3 and 4 display these curves for two different values of the coefficient η of the regularizer. (Notice the logarithmic scale on the vertical axes in these figures.)

The curves of constant q_0 have two branches now. For a given q_0 the lower branch lies mostly or partly in the previously feasible region, the upper branch lies outside,

J. Stat. Mech. (2019) 013402

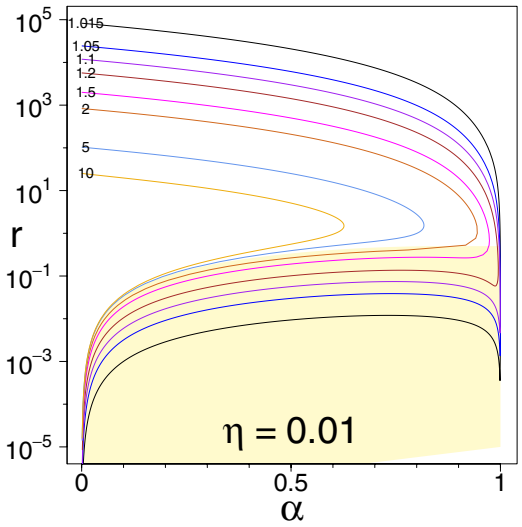

Figure 3. Contour plot for fixed values of $\sqrt{q_0}$ on the $\alpha - r$ plane at $\eta = 0.01$.

above it. Along the lower branch the value of the ratio r is small, which means we have very large samples with respect to the dimension: this is the data-dominated regime. We can also see that when the data dominate, the dependence on the regularizer is weak: the set of curves inside the yellow region is quite similar in the two figures, even though the regularizer has been increased 5-fold from figure 3 to 4. Following the curve corresponding to a given value of q_0, say the black one, we see that at the beginning it is increasing with α, in the vicinity of $\alpha = 1$ it starts to decline, then it sharply turns around and shoots up steeply, leaving the feasible region and increasing with decreasing α. Along this upper branch the ratio r is not small any more. We do not have large samples here, in fact, the situation is just the opposite: the dimension N becomes larger than the size T of the samples. Clearly, in this regime the regularizer dominates and the data play only a minor role: this is the bias-dominated regime. It is interesting to note the sudden turn over of the constant q_0 curves in the vicinity of $\alpha = 1$. Such a sharp feature would be extremely hard to discover if we wanted to solve the original optimization problem by numerical simulations: the simulation would jump over to the upper branch before we could observe the sharp dip and the identification of the boundary between the data-dominated and the bias-dominated regimes would be hard. This is even more so for real life data which are inevitably noisy.

An important point in regularization is the correct choice of the parameter η. When data come from real observations, and the size of the sample (or the number of samples) is limited by time and/or cost considerations, the standard procedure is cross validation [27], i.e. using a part of the data to infer the value of η and checking the correctness of this choice on the other part. In the present analytical approach we have the luxury of infinitely many samples to average over, so we can obtain the value of the coefficient of regularization by demanding a given relative error (that is a given q_0) for a given confidence limit α and given aspect ratio $r = N/T$. Figure 5 displays the plot of the given estimation error curves on the $r - \eta$ plane for the specific value of $\alpha = 0.975$ demanded by the new market risk regulation [1], and relative errors of 1%, 5% and 10%, respectively. It shows the change-over between the data-dominated resp. bias-dominated regimes very clearly. For a given value of r the corresponding value

J. Stat. Mech. (2019) 013402

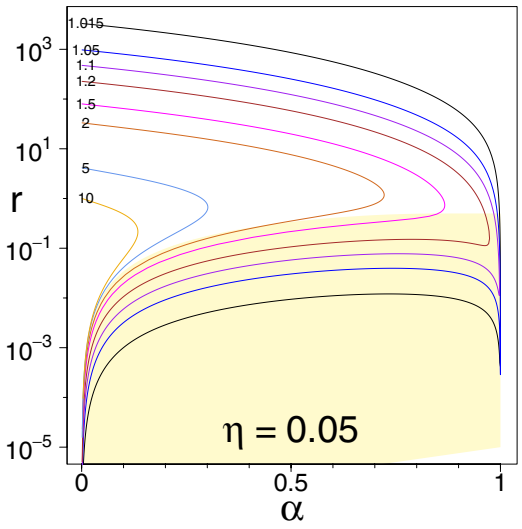

Figure 4. Contour plot for fixed values of $\sqrt{q_0}$ on the $\alpha - r$ plane at $\eta = 0.05$.

of η can be read off from the curves. If r is small (i.e. the sample is large with respect to the dimension) the curves with the prescribed values of relative error are almost horizontal. This means that when we have sufficient data the value of the regularizer is more or less immaterial: within reasonable limits we can choose any coefficient for the regularizer, it will not change the precision of the estimate, because in this situation the data will determine the optimum. Conversely, when the data are insufficient (r is not small, or it is even beyond the feasible region), the value of η necessary to enforce a given relative error strongly depends on r. In this region, however, we need a smaller and smaller η to find the same relative error, because here the data almost do not matter and even a small regularizer will establish the optimum. The transition between these two regimes takes place around the points where the curves turn back. This happens still inside the feasible region, and the width of this range is rather small: from the r value corresponding to $\eta = 0$ to the one where the curves turn around the increase of r always remains within a factor of about 4.

Let us take a closer look at that part of the previous figure where the curves turn around and r starts to increase. Figure 6(a) shows this region in higher resolution. For a given, small, value of the estimation error (such as 1% or 2%), r grows by a factor of about four by the time we reach the elbow of the curves (at rather large η values). This means that for a given sample size T the regularization allows us to consider a four times larger portfolio without increasing the estimation error. Conversely, for a given value of N the regularizer allows the use of four times shorter time series without compromising the quality of the estimate. Of course, the growth of r could be followed beyond the elbow, up to higher values along the constant q_0 curves, but it must be clear that these sections of the curves correspond to a situation where the estimate is mostly or entirely determined by the regularizer. This is also shown by the fact that the curves of given estimation error strongly lean backwards to the vertical axis: where the dimension is high and the data few even a weak regularizer can stabilize the estimate, but it will then have nothing to do with the information coming from the time series.

J. Stat. Mech. (2019) 013402

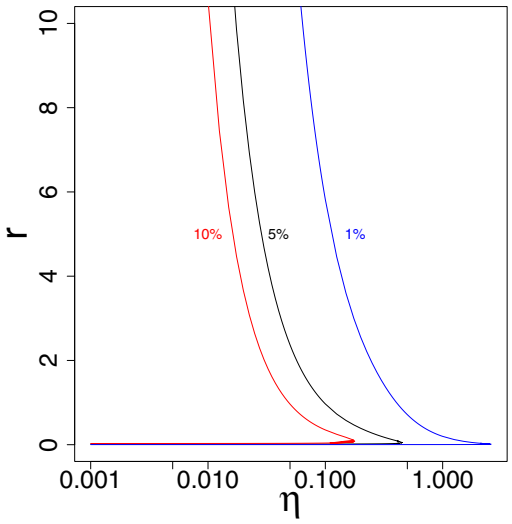

Figure 5. The overall behavior of the contour lines of fixed estimation error (fixed q_0) on the $r = N/T - \eta$ plane, for a given value of the confidence limit $\alpha = 0.975$ and for three different values (1%, 5% and 10%) of the relative estimation error. The data-dominated and bias-dominated regions correspond to the two branches of these curves: in the range of small r's the curves depend on the strength of the regularizer very weakly, while for r's in the vicinity of the phase boundary, and even more for large r's high in the originally unfeasible region, the fixed estimation error curves display a strong dependence on η.

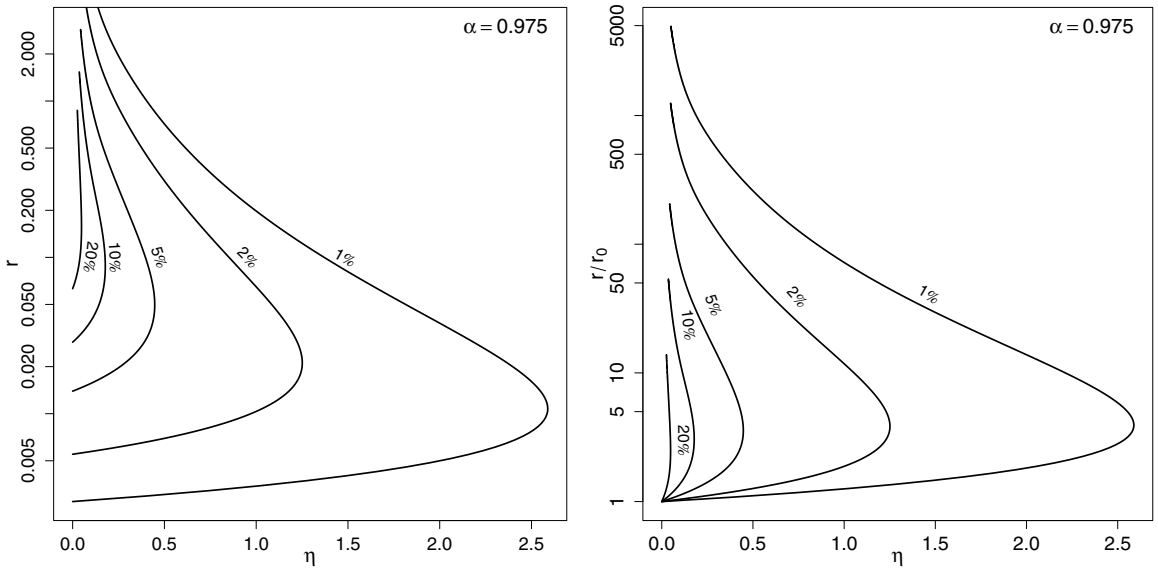

Figure 6. (a) The previous figure in higher resolution (left). (b) The same as the left figure, but the $r(\eta)$ curves normalized by their initial values r_0 corresponding to $\eta = 0$ (right). It can be seen that the gain in r is about a factor 4.

A gain of a factor four in the allowed region in r could be regarded as very satisfactory, were it not for the fact that the initial ($\eta = 0$) value of r along the small estimation error curves is so small that it remains small even after a multiplication by 4.

J. Stat. Mech. (2019) 013402

If we inspect another curve, corresponding to a larger estimation error (say, 5%), we can see that it turns back for a much smaller η, but the relative increase of r up to the elbow is still about a factor 4. We can also see that beyond this point the curves very quickly reach the region where the regularizer dominates.

Figure 6(b) displays the same curves as in figure 6(a), but this time they are normalized by their vales at $\eta = 0$, so that they show the gain in r due to the regularizer.

4. Conclusion

We have considered the problem of optimizing ES in the presence of an ℓ_2 regularizer for the case of uncorrelated Gaussian returns. The regularizer takes care of the large sample fluctuations and eliminates the phase transition that would be present in the problem without regularization. Deep inside the feasible region, where we have a large amount of data relative to the dimension, the size of the sample needed for a given level of relative estimation error is basically constant, largely independent of the regularizer. In the opposite case, for sample sizes comparable to or small relative to the dimension, the regularizer dominates the optimization and suppresses the data. The transition between the the data-dominated regime and the regularizer-dominated one is rather narrow. It is in this transition region where we can meaningfully speak about a trade-off between fluctuation and bias, otherwise one or the other dominates the estimation. The identification of this transitional zone is easy within the present scheme, where we could carry out the optimization analytically: the transitional zone is the small region where the curves in figure 5 sharply turn back, but still remain inside the originally feasible region. In real life, where the size of the samples can rarely be changed at will and where all kinds of external noise (other than that coming from the sample fluctuations) may be present, the distinction between the region where the data dominates and where the bias reigns may be much less clear, and one may not be sure where the transition takes place between them. Below this transition there is not much point in using regularization, because the data themselves are sufficient to provide a stable and reliable estimate. Above the transition zone it is almost meaningless to talk about the observed data, because they are crowded out by the bias. The identification of the relatively narrow transition zone between these two extremes and the gain of a factor four below the transition are the main results of this paper.

It is important to stress that the results we presented in this paper have been obtained under the simplifying assumption that asset returns are i.i.d. Gaussian variables. This was done partly for consistency with previous work [13–17], but, more importantly, for mathematical convenience, as under this assumption it is relatively easy to provide an analytical characterization of the problem. With respect to real market data the assumption is unrealistic for two reasons: first, the distribution of returns is known to be different from a normal distribution, it displays fat tails, although with the general acceleration of trading the convergence to the Gaussian is becoming faster, especially on the time scales relevant for portfolio rebalancing. Second, the market is characterized by a non-trivial structure of correlations between assets, with modes representing the movement of the market as a whole and of the different market sectors.

J. Stat. Mech. (2019) 013402

Relaxing the assumption of normality cannot be easily accommodated in the replica calculation, however it is not expected to change dramatically the results discussed in the paper. This was for instance shown in [24], where the non-regularized problem was studied by means of numerical simulations for the case of uncorrelated returns distributed according to a student's t-distribution and it was shown that the critical line was the same as for the case of Gaussian variables. Introducing correlations between assets would also not change the qualitative features discussed in the paper: the regularizer would still prevent the divergence of the estimation error, and we would still observe a data-dominated region, where the effect of regularization is negligible, and a region where the solution of the optimization problem is dominated by the regularizer, where data play little role. However, the gain in performance due to the regularizer and the width of the transition region may in general depend on the details of the correlation structure. This will be the subject of future work.

Finally, it is important to realize that the cause of the relatively narrow transition region we identify is the same as that of the strong fluctuations, namely the unbounded loss function. Expected shortfall is not the only risk measure to have this deficiency: all the downside risk measures have it, including VaR. The preference for downside risk measures is explained by the fact that investors (and regulators) are not afraid of big gains, only of big losses. Perhaps they should be. Refusing to acknowledge the risk in improbably large gains is a Ponzi scheme mentality. Downside risk measures embody this mentality. As a part of regulation, however, they acquire an air of undeserved respectability, at which point the associated technical issues become components of systemic risk.

Acknowledgments

We are obliged to R Kondor, M Marsili and S Still for valuable discussions. FC acknowledges support of the Economic and Social Research Council (ESRC) in funding the Systemic Risk Centre (ES/K002309/1). IK is grateful for the hospitality extended to him at the Computer Science Department of the University College of London where this paper was written up.

References

[1] Basel Committee on Banking Supervision 2016 Minimum capital requirements for market risk (Basel: Bank for International Settlements)
[2] Jorion P 1986 Bayes–Stein estimation for portfolio analysis *J. Financ. Quant. Anal.* **21** 279–92
[3] Frost P A and Savarino J E 1986 An empirical Bayes approach to efficient portfolio selection *J. Financ. Quant. Anal.* **21** 293–305
[4] Ledoit O and Wolf M 2003 Improved estimation of the covariance matrix of stock returns with an application to portfolio selection *J. Empir. Finance* **10** 603–21
[5] Ledoit O and Wolf M 2004 A well-conditioned estimator for large-dimensional covariance matrices *J. Multivar. Anal.* **88** 365–411
[6] Ledoit O and Wolf M 2004 Honey, I shrunk the sample covariance matrix *J. Portfolio Manage.* **31** 110
[7] Golosnoy V and Okhrin Y 2007 Multivariate shrinkage for optimal portfolio weights *Eur. J. Finance* **13** 441–58
[8] Brodie J, Daubechies I, De Mol C, Giannone D and Loris I 2009 Sparse and stable Markowitz portfolios *Proc. Natl Acad. Sci.* **106** 12267–72

J. Stat. Mech. (2019) 013402

[9] Laloux L, Cizeau P, Bouchaud J-P and Potters M 1999 Noise dressing of financial correlation matrices *Phys. Rev. Lett.* **83** 1467–70
[10] Laloux L, Cizeau P, Bouchaud J-P and Potters M 2000 Random matrix theory and financial correlations *Int. J. Theor. Appl. Finance* **3** 391
[11] Plerou V, Gopikrishnanm P, Rosenow B, Amaral L A N and Stanley H E 1999 Universal and non-universal properties of cross-correlations in financial time series *Phys. Rev. Lett.* **83** 1471
[12] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N, Guhr T and Stanley H E 2000 A random matrix approach to cross-correlations in financial time-series *Phys. Rev.* E **65** 066136
[13] Ciliberti S, Kondor I and Mézard M 2007 On the feasibility of portfolio optimization under expected shortfall *Quant. Finance* **7** 389–96
[14] Ciliberti S and Mézard M 2007 Risk minimization through portfolio replication *Eur. Phys. J.* B **57** 175–80
[15] Varga-Haszonits I and Kondor I 2008 The instability of downside risk measures *J. Stat. Mech.* P12007
[16] Caccioli F, Still S, Marsili M and Kondor I 2013 Optimal liquidation strategies regularize portfolio selection *Eur. J. Finance* **19** 554–71
[17] Caccioli F, Kondor I, Marsili M and Still S 2016 Liquidity risk and instabilities in portfolio optimization *Int. J. Theor. Appl. Finance* **19** 1650035
[18] Takeda A and Sugiyama M 2008 ν-support vector machine as conditional value-at-risk minimization *Proc. of the 25th Int. Conf. on Machine Learning* (ACM) pp 1056–63
[19] Still S and Kondor I 2010 Regularizing portfolio optimization *New J. Phys.* **12** 075034
[20] Pérez-Cruz F, Weston J, Herrmann D J L and Scholkopf B 2003 Extension of the ν-svm range for classification *Nato Sci. Ser. III Comput. Syst. Sci.* **190** 179–96
[21] Bouchaud J-P and Potters M 2003 *Theory of Financial Risk and Derivative Pricing* (Cambridge: Cambridge University Press)
[22] Gábor A and Kondor I 1999 Portfolios with nonlinear constraints and spin glasses *Physica* A **274** 222–8
[23] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (*World Scientific Lecture Notes in Physics* vol 9) (Singapore: World Scientific)
[24] Caccioli F, Kondor I and Papp G 2015 Portfolio optimization under expected shortfall: contour maps of estimation error *Quantitative Finance* **18** 1295–313
[25] Rockafellar R T and Uryasev S 2000 Optimization of conditional value-at-risk *J. Risk* **2** 21–41
[26] Kondor I and Varga-Haszonits I 2010 Instability of portfolio optimization under coherent risk measures *Adv. Complex Syst.* **13** 425–37
[27] Hastie T, Tibshirani R and Friedman J 2008 *The Elements of Statistical Learning, Data Mining, Inference, and Prediction (Springer Series in Statistics)* 2nd edn (Berlin: Springer)

J. Stat. Mech. (2019) 013402