entropy

MDPI

Article

Optimizing Expected Shortfall under an ℓ_1 Constraint—An Analytic Approach

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

1 Institute for Physics, Eötvös Loránd University, 1117 Budapest, Hungary; pg@ludens.elte.hu
2 Parmenides Foundation, 82049 Pullach, Germany
3 London Mathematical Laboratory, London W6 8RH, UK; f.caccioli@ucl.ac.uk
4 Complexity Science Hub, Vienna 1080, Austria
5 Department of Computer Science, University College London, London WC1E 6BT, UK
6 Systemic Risk Centre, London School of Economics and Political Sciences, London WC2A 2AE, UK
* Correspondence: i.kondor@lml.org.uk

Abstract: Expected Shortfall (ES), the average loss above a high quantile, is the current financial regulatory market risk measure. Its estimation and optimization are highly unstable against sample fluctuations and become impossible above a critical ratio $r = N/T$, where N is the number of different assets in the portfolio, and T is the length of the available time series. The critical ratio depends on the confidence level α, which means we have a line of critical points on the $\alpha - r$ plane. The large fluctuations in the estimation of ES can be attenuated by the application of regularizers. In this paper, we calculate ES analytically under an ℓ_1 regularizer by the method of replicas borrowed from the statistical physics of random systems. The ban on short selling, i.e., a constraint rendering all the portfolio weights non-negative, is a special case of an asymmetric ℓ_1 regularizer. Results are presented for the out-of-sample and the in-sample estimator of the regularized ES, the estimation error, the distribution of the optimal portfolio weights, and the density of the assets eliminated from the portfolio by the regularizer. It is shown that the no-short constraint acts as a high volatility cutoff, in the sense that it sets the weights of the high volatility elements to zero with higher probability than those of the low volatility items. This cutoff renormalizes the aspect ratio $r = N/T$, thereby extending the range of the feasibility of optimization. We find that there is a nontrivial mapping between the regularized and unregularized problems, corresponding to a renormalization of the order parameters.

Keywords: portfolio optimization; regularization; renormalization

check for updates

Citation: Papp, G.; Kondor, I.; Caccioli, F. Optimizing Expected Shortfall under an ℓ_1 Constraint—An Analytic Approach. *Entropy* **2021**, *23*, 523. https://doi.org/10.3390/e23050523

Academic Editors: Ryszard Kutner and Geert Verdoolaege

Received: 9 March 2021
Accepted: 22 April 2021
Published: 24 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

A risk measure is a functional on the probability distribution of the fluctuating returns of a security or a portfolio. Since it is impossible to condense all the information in a probability distribution into a single number, there is no unique way to choose the "best" risk measure. In Markowitz's ground breaking portfolio selection theory [1], with the assumption of Gaussian distributed returns, variance offered itself as the natural risk measure. The crises of the late eighties and early nineties led both the industry and regulators to realize that the most dangerous risk lurked in the asymptotically far tail of the return distribution. To grasp this risk, a high quantile of the profit and loss distribution called Value at Risk (VaR) was introduced by J.P. Morgan [2]. For a certain period, VaR became a kind of industry standard, and it was embraced by international financial regulation as the official risk measure in 1996 [3]. Value at Risk is a threshold which losses only exceed with a small probability (such as, e.g., 0.05 or 0.01), corresponding to a confidence level of $\alpha = 0.95$, resp. 0.99. (In this context, it is

customary to regard losses as positive and profits as negative). As a quantile, VaR is not sensitive to the distribution of losses above the confidence level and is not subadditive when two portfolios are combined. This triggered a search for alternatives and led Artzner et al. [4] to formulate a set of axioms that any coherent risk measure should satisfy. The simplest and most intuitive of these coherent measures is the Expected Shortfall (ES) [5,6]. ES is essentially the expected loss above a high quantile that can be chosen to be the VaR itself. After a long debate about the relative merits and drawbacks of ES, whose details are not pertinent to our present study, regulators adopted ES as the current official market risk measure to be used to assess the financial health of banks and determine the capital charge they are required to hold against their risks. The regulators and the industry settled on a confidence level of $\alpha = 0.975$ [7].

ES is mainly designed to be a diagnostic tool. At the same time, it is also a constraint that banks have to respect when considering the composition of their portfolios. It is then in their best interest to optimize ES, in order to keep their capital charge as low as possible. However, the optimization of ES is fraught with problems of estimation error, which is quite natural if one considers that the number of different items N in a bank's portfolio can be very large, whereas the number of observations (the length of the available time series T) is always limited. In addition, at the regulatory confidence level, one has to throw away 97.5% of the data. Moreover, the estimation error increases with the ratio $r = N/T$ and at a critical value of r, it actually diverges, growing beyond any limit. As shown in [8], the instability of the optimization of ES (as well as all the coherent risk measures) follows directly from the coherence axioms [4].

The divergence of ES is the signature of a phase transition. The critical r for ES is smaller or equal to $1/2$, its value depending on the confidence level α. For ES, there is then a line of critical points, a phase diagram, on the $r - \alpha$ plane. A part of this phase diagram has been traced out by numerical simulations in [9], while the full phase diagram has been determined by analytical calculations by Ciliberti et al. [10]. Going beyond merely determining the phase diagram, a detailed study of the estimation error and other relevant quantities has been performed inside the whole feasibility region in [11,12], and it was shown that, due to the nontrivial behavior of the contour lines of constant estimation error, especially in the vicinity of $\alpha = 1$, the number of data necessary to have a reasonably low estimation error was way above any T available in practice.

Because of the large sample fluctuations of ES, its optimization constitutes a problem in high dimensional statistics [13]. A standard tool to tame these large fluctuations is to introduce regularizers, which penalize large excursions. Although the introduction of these penalties may seem an arbitrary statistical trick coming from outside of finance, it was shown in [14] that these regularizers express liquidity considerations, and take into account, already at the construction of the portfolio, the expected market impact of a future liquidation. The regularizers are usually chosen to be some constraints on the norm of the portfolio weights. In [15], we studied the effect of an ℓ_2 regularizer on ES and found that ℓ_2 obviously suppresses the instability and, for sufficiently small r and with a strong enough regularizer, it extends the range where the estimation error is reasonably small by a factor of about 4.

It is interesting to see how an ℓ_1 regularizer works with ES. (The importance of studying the effect of various regularizers in combination with the different risk measures was emphasized by [16]). The regularizer ℓ_1 is known to produce sparse solutions, which means that in order to rein in large fluctuations, it eliminates some of the securities from the portfolio. This obviously contradicts the principle of diversification, but considerations of transaction costs or the technical difficulties of managing large portfolios may make it desirable to remove the most volatile items from the portfolio, and this is precisely what a no-short constraint tends to do.

It has been known for 20 years now that the optimization of ES can be translated into a linear programming problem [17]. Accordingly, as it has been realized in [18], the piece-wise linear ℓ_1 with an infinite slope corresponding to an infinite penalty on short selling can prevent the instability of ES. The purpose of this paper is to determine the effect of ℓ_1-regularization on the phase diagram and also on the behavior of the various quantities of interest inside the region where the optimization of ES is feasible and meaningful. (We will see that as a result of regularization new characteristic lines appear on the $r - \alpha$ plane, beyond which the optimization of ES is still mathematically feasible, but the results become meaningless, as they correspond to negative risk.) In [12], a detailed analytical investigation of the behavior of the estimation error, the in-sample cost, the sensitivity to small changes in the composition of the portfolio, and the distribution of optimal weights were carried out in the non-regularized case. Here, we derive the same quantities for an ℓ_1-regularized ES, including the special case where short selling is banned, that is when the portfolio weights are constrained to be non-negative. The density of the items eliminated from the portfolio, to be referred to as the "condensate" in the following, is also determined. The most striking result of the present study is that the regularized solution can be mapped back onto the unregularized one. We are not aware of a similarly tight relationship between a regularized and an unregularized problem, not only in a finance context, but neither in the general context of machine learning.

2. Method and Preliminaries

If the true probability distribution of returns were known, it would be easy to calculate the true value of Expected Shortfall and the optimal portfolio weights. However, the true distribution of returns is unknown, therefore one has to rely on finite samples of empirical data. This means one observes N time series of length T and estimates the optimal weights and ES on the basis of this information. It is clear that the weights and ES so obtained will deviate from their "true" values. (The latter would be obtained in an infinitely long stationary sample.) The deviation of the estimated values will be the stronger the shorter the length T and the larger the dimension N. Performing this measurement on different samples one would obtain different estimates: there is a distribution of ES and of the optimal weights over the samples. In a real market, one cannot repeat such an experiment multiple times. Instead, one has to squeeze out as much information as possible from a single sample of limited size. There are well-known numerical methods for this, like cross-validation or bootstrap [19]. In contrast, in the present work we aim to obtain *analytic* results. In order to mimic empirical sampling, we choose a simple data generating process, such as a multivariate Gaussian. The true value of ES is easy to obtain for this case, which provides a standard to measure finite sample deviations from. Then we determine ES for a large number of random samples of length T drawn from this underlying distribution, average it over the random samples and finally compare this average to its true value. This procedure will give us an idea about how large the estimation error is for a given dimension N, sample size T, and confidence level α, under the idealized conditions of stationarity and Gaussian fluctuations, and how much it will be reduced when we apply an ℓ_1 regularizer of a given strength. It is reasonable to assume that the estimation error obtained under these idealized circumstances will be a lower bound to the estimation error for real-life processes.

Now we wish to implement this program via analytic calculations. The averaging over the random samples just described is analogous to the averaging over the random realization of disorder in the statistical physics of random systems, which enables us to borrow methods from that field, in particular the replica method [20]. It assumes that both N and T are large, with their ratio $r = N/T$ kept finite (thermodynamic or Kolmogorov limit). A small value of r corresponds to the classical setup in statistics where one has a large number of observations relative to the dimension. Estimates in this case are sharp and close to their true values.

In contrast, when r is of order unity, or larger, we are in the high dimensional limit where fluctuations are large. It is here that the regularizer becomes important.

In the usual application of ℓ_1 in finite dimensional numerical studies, the regularizer eliminates the dimensions one by one, in a stepwise manner, as the strength of the regularizer is increasing. In our present work, the large N, T limit and the averaging over infinitely many samples result in a continuous dependence of the "condensate" density (the relative number N_0/N of the dimensions eliminated by ℓ_1) on the aspect ratio r, the confidence level α, and the strength of ℓ_1. In a study of ℓ_1-regularized variance [21], we found that the stepwise increase of the density of eliminated weights in a numerical experiment nicely follows the continuous curve obtained analytically. It is obvious that the situation is similar in the case of ES, but we have also confirmed this by numerical simulations.

For the sake of simplicity, we will also assume that the returns are independent, that is the true covariance matrix is diagonal. This is not an innocent assumption: it will be seen, for example, that the maximum degree of sparsity that ℓ_1 can achieve in this scheme is one half of the total number of dimensions, whereas for correlated returns the maximum sparsity can be either larger or smaller than $1/2$, according to whether correlations are predominantly positive or negative. Combining ℓ_1 with a non-diagonal covariance matrix poses additional technical difficulties that we wish to avoid in the present account. However, we do allow the diagonal elements σ_i of the covariance matrix to be different from each other.

As a further simplification, we do not impose any other constraint on the optimization of ES beside the budget constraint and the ℓ_1 regularizer. In particular, we do not set a constraint on the expected return, and seek the global minimum of the regularized ES. This is in line with a number of studies, [22–24] among others, which focus on the global minimum in the problem of variance optimization, because of the extremely noisy estimates of the expected return. Furthermore, the global minimum is precisely what one needs in minimizing tracking-errors, that is, when trying to follow, say, a market index as closely as possible [23].

The replica method used below have already been applied with minor variations to various portfolio optimization problems in a number of papers [10–12,14,18,21,25–28], where the replica derivation of the main formulae were repeatedly explained, so we do not need to go through that exercise again here. Then the natural starting point for our present work is the detailed study of the behavior of ES *without* regularization in [12]. The argument there leads to a relationship between ES and an effective cost or free energy per asset f as follows:

$$\mathrm{ES} = \frac{fr}{1-\alpha}\,. \tag{1}$$

The free energy f itself is given by the minimum of a functional depending on six order parameters

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

where

$$V(w,z,\sigma) = \hat{\Delta}\sigma^2 w^2 - \lambda w - zw\sigma\sqrt{-2\hat{q}_0} + \eta^+\theta(w)w - \eta^-\theta(-w)w \tag{3}$$

and the double average $\langle\ldots\rangle_{\sigma,z}$ means

$$\int_0^\infty \mathrm{d}\sigma\, \frac{1}{N}\sum_i \delta(\sigma-\sigma_i)\int_{-\infty}^{\infty}\frac{\mathrm{d}z}{\sqrt{2\pi}}e^{-z^2/2}\ldots \tag{4}$$

Finally, the function g in the integral in (2) is defined as

$$g(x) = \begin{cases} 0, & x \geq 0 \\ x^2, & -1 \leq x \leq 0 \\ -2x-1, & x < -1 \end{cases} . \tag{5}$$

The differences with respect to the setup in [12] are the following: a trivial change of notation (τ there is $1/r$ here); the variable σ has been introduced in (3), which together with the recipe (4) allows us to consider assets with different volatilities σ_i; and the regularizer has been built into the effective potential (3). Note that the ℓ_1 in (3) is asymmetric in order to allow us to penalize long and short positions separately. The usual ℓ_1 corresponds to $\eta^+ = \eta^-$, the ban on short selling to $\eta^- \to \infty$. We will also use the arrangement where there is a finite penalty η^- on short positions and none on long ones $\eta^+ = 0$.

A note on signs: for consistency, the order parameters λ, Δ, q_0, and $\hat{\Delta}$ must be positive, $\hat{q}_0$ negative, and ϵ can be of either sign. Furthermore, λ must be larger or equal to the right slope of the regularizer: $\lambda \geq \eta^+$.

Before setting out to derive the stationarity conditions that determine the optimal value of the free energy and thence of ES, we spell out the meaning of the order parameters. The first of these is the Lagrange multiplier λ that enforces the budget constraint:

$$\sum_{i=1}^{N} w_i = N. \tag{6}$$

Note that the sum of portfolio weights is set to N here, instead of the usual 1. This is to keep the weights of order unity in the large N limit.

Because of the relationship between λ and the budget constraint, λ can be thought of as a kind of chemical potential. It is an important quantity, because, as we shall see later, its value at the stationary point is equal to the free energy, hence directly related to the optimal value of ES. In [12], we argued that this optimal value of ES is, in fact, the in-sample estimate of Expected Shortfall. According to (1), ES is proportional to the product fr, which means f, and hence λ too, must be inversely proportional to r when $r = N/T \to 0$, because ES is certainly finite in this limit: a finite N and $T \to \infty$ corresponds to the case of having complete information. This spurious divergence of f and λ is an artifact, due to our having absorbed a factor $1/r$ in their definition. This is explained purely by convenience: we wish to keep as close to the convention in [12] as possible. The opposite limit, when $\lambda - \eta^+$ vanishes, is another important point: it signals the instability of the portfolio, and the onset of the phase transition.

The next order parameter, ϵ, was suggested by [17] as a proxy for Value at Risk. Indeed, in the limit $r \to 0$ where we know the true distribution of returns, ϵ will be seen to be equal to the known value of VaR for a Gaussian.

The third order parameter, q_0, is of central importance: According to [12], the ratio of the out-of-sample estimate ES_{out} and its true value $\mathrm{ES}^{(0)}$ is given by the square root of q_0. For the case of different σ_is considered here, q_0 has to be amended by a factor depending on the structure of the portfolio [21] as

$$\tilde{q}_0 = q_0 \frac{1}{N} \sum_i \frac{1}{\sigma_i^2}. \tag{7}$$

Then the ratio of the estimated and true ES will be

$$\frac{\mathrm{ES}_{out}}{\mathrm{ES}^{(0)}} = \sqrt{\tilde{q}_0} \tag{8}$$

that is the relative estimation error is $\sqrt{\tilde{q}_0} - 1$.

The fourth order parameter, Δ, measures the sensitivity to a small shift in the returns.

The remaining two order parameters, $\hat{q}_0$ and $\hat{\Delta}$, are auxiliary variables that do not have an obvious meaning, they enter the picture through the replica formalism, and can be eliminated once the stationarity conditions have been established. The stationarity or saddle point conditions are derived by taking the derivative of the free energy with respect to the order parameters and setting them to zero. They will be written up in the next Section.

3. Results

First, we are going to spell out the saddle point conditions in full detail and reduce them to special cases later.

Let us bring the integral in (2) to a more convenient form by integrating by parts:

$$I = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \mathrm{d}s\, e^{-s^2} g\left(\frac{\epsilon}{\Delta} + s\sqrt{\frac{2q_0}{\Delta^2}}\right) = \frac{2q_0}{\Delta^2}\left[W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right] - 1 - 2\frac{\epsilon}{\Delta}\,. \tag{9}$$

With this identity, the free energy becomes

$$f = \lambda - \frac{\alpha\epsilon}{r} - \Delta\hat{q}_0 - \hat{\Delta}q_0 - \frac{\Delta}{2r} + \frac{q_0}{r\Delta}\left[W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right] + \langle \min V\rangle_{\sigma,z}\,. \tag{10}$$

The function W in the above formulae, together with two related functions Φ and Ψ, will frequently appear in the following; they are integrals of the Gaussian $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$:

$$\Phi(x) = \int_{-\infty}^{x} \mathrm{d}t\, \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \tag{11}$$

$$\Psi(x) = \int_{-\infty}^{x} \mathrm{d}t\, \Phi(t) \tag{12}$$

$$W(x) = \int_{-\infty}^{x} \mathrm{d}t\, \Psi(t)\,. \tag{13}$$

Now we evaluate the minimum of V in (3) and denote the "representative weight" where this minimum is located by w^*. It works out to be

$$w^* = \frac{\lambda + \sigma z\sqrt{-2\hat{q}_0} - \eta^+\Theta(w^*) + \eta^-\Theta(-w^*)}{2\sigma^2\hat{\Delta}}\,, \tag{14}$$

or

$$w^* = \begin{cases} \frac{\lambda+\sigma z\sqrt{-2\hat{q}_0}-\eta^+}{2\sigma^2\hat{\Delta}}, \text{ if } z \geq \frac{\eta^+-\lambda}{\sigma\sqrt{-2\hat{q}_0}} \\ 0, \text{ if } -\frac{\lambda+\eta^-}{\sigma\sqrt{-2\hat{q}_0}} < z < \frac{\eta^+-\lambda}{\sigma\sqrt{-2\hat{q}_0}} \\ \frac{\lambda+\sigma z\sqrt{-2\hat{q}_0}+\eta^-}{2\sigma^2\hat{\Delta}}, \text{ if } z \leq -\frac{\lambda+\eta^-}{\sigma\sqrt{-2\hat{q}_0}}\,. \end{cases} \tag{15}$$

With this and (4), one can calculate V^*, the value of V at the minimum, and perform the double averaging to obtain

$$\langle V^*\rangle_{\sigma,z} = \frac{\hat{q}_0}{\hat{\Delta}}\frac{1}{N}\sum_i \left[W\left(\frac{\lambda-\eta^+}{\sigma_i\sqrt{-2\hat{q}_0}}\right) + W\left(-\frac{\lambda+\eta^-}{\sigma_i\sqrt{-2\hat{q}_0}}\right)\right]. \tag{16}$$

Then, the fully explicit form of the free energy becomes

$$\begin{aligned} f &= \lambda - \frac{\alpha\epsilon}{r} - \Delta\hat{q}_0 - \hat{\Delta}q_0 - \frac{\Delta}{2r} + \frac{q_0}{r\Delta}\left[W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right] \\ &+ \frac{\hat{q}_0}{\hat{\Delta}}\frac{1}{N}\sum_i\left[W\left(\frac{\lambda-\eta^+}{\sigma_i\sqrt{-2\hat{q}_0}}\right) + W\left(-\frac{\lambda+\eta^-}{\sigma_i\sqrt{-2\hat{q}_0}}\right)\right]. \end{aligned} \tag{17}$$

It is now straightforward to take the derivatives of f with respect to the order parameters and derive the stationary conditions.

From $\partial f/\partial\lambda = 0$, it follows that

$$1 = \frac{\sqrt{-2\hat{q}_0}}{2\hat{\Delta}}\frac{1}{N}\sum_i\frac{1}{\sigma_i}\left[\Psi\left(\frac{\lambda-\eta^+}{\sigma_i\sqrt{-2\hat{q}_0}}\right) - \Psi\left(-\frac{\lambda+\eta^-}{\sigma_i\sqrt{-2\hat{q}_0}}\right)\right]. \tag{18}$$

The derivative with respect to $\hat{q}_0$ yields

$$2\Delta\hat{\Delta} = \frac{1}{N}\sum_i\left[\Phi\left(\frac{\lambda-\eta^+}{\sigma_i\sqrt{-2\hat{q}_0}}\right) + \Phi\left(-\frac{\lambda+\eta^-}{\sigma_i\sqrt{-2\hat{q}_0}}\right)\right]. \tag{19}$$

From the derivative with respect to $\hat{\Delta}$, we get

$$q_0 = -\frac{\hat{q}_0}{\hat{\Delta}^2}\frac{1}{N}\sum_i\left[W\left(\frac{\lambda-\eta^+}{\sigma_i\sqrt{-2\hat{q}_0}}\right) + W\left(-\frac{\lambda+\eta^-}{\sigma_i\sqrt{-2\hat{q}_0}}\right)\right]. \tag{20}$$

As mentioned before, q_0 determines the out-of-sample estimate for ES and the estimation error.

The derivative with respect to q_0 leads to

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

where use has been made of the identity

$$W(x) = \frac{1}{2}x\Psi(x) + \frac{1}{2}\Phi(x)\,. \tag{22}$$

The condition for the derivative with respect to ϵ to vanish is

$$\alpha = \frac{\sqrt{q_0}}{\Delta}\left[\Psi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - \Psi\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right]. \tag{23}$$

The derivation of the last equation takes a little more effort. Let us go back to the free energy in (2) and take the derivative with respect to Δ. Noticing that $\langle V\rangle_{\sigma,z}$ does not depend on Δ, and using the integral given in (9), we have

$$\frac{\partial f}{\partial\Delta} = -\hat{q}_0 + \frac{1}{2r}I + \frac{\Delta}{2r}\frac{\partial I}{\partial\Delta} = 0 \tag{24}$$

valid at the stationary point. From here we find

$$\frac{1}{2r}I_{st} = \hat{q}_0 + \frac{2q_0}{r\Delta^2}\left[W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right] - \frac{\epsilon}{r\Delta} - \frac{\sqrt{q_0}}{r\Delta}\Psi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right), \tag{25}$$

where (9) was used again and we denoted by I_{st} the integral I evaluated at the stationary point. Now we apply the identity (22) and the stationary conditions (23), (21) to arrive at

$$\frac{1}{2r} I_{st} = \hat{q}_0 + \frac{2q_0\hat{\Delta}}{\Delta} - (1-\alpha)\frac{\epsilon}{r\Delta}\,, \tag{26}$$

which, combined with (9), finally leads to

$$\hat{q}_0 + \frac{2q_0\hat{\Delta}}{\Delta} + \alpha\frac{\epsilon}{r\Delta} + \frac{1}{2r} - \frac{q_0}{r\Delta^2}\left[W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right] = 0\,. \tag{27}$$

The Equations (18)–(23) and (27) constitute the system of equations for the six order parameters. These equations are valid both for the regularized and (setting $\eta^+ = \eta^- = 0$) for the unregularized cases.

Let us now work out the relationship between the free energy and the chemical potential. Comparing (16) and (20), we see that $\langle V^* \rangle_{\sigma,z} = -q_0\hat{\Delta}$, which with (10) and (27), results in the simple formula

$$f = \lambda \tag{28}$$

at the stationary point, as we anticipated before. In [12], we argued that the stationary value of f determines the in-sample estimate of ES through (1).

The last object to determine is the distribution of weights:

$$p(w) = \langle \delta(w - w^*) \rangle_{\sigma,z}\,. \tag{29}$$

With (14), we find

$$p(w) = n_0\delta(w) + \frac{1}{N}\sum_i \frac{1}{\sigma_w^{(i)}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{w - w_i^+}{\sigma_w^{(i)}}\right)^2\right)\theta(w) \tag{30}$$

$$+ \frac{1}{N}\sum_i \frac{1}{\sigma_w^{(i)}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{w - w_i^-}{\sigma_w^{(i)}}\right)^2\right)\theta(-w)\,, \tag{31}$$

where $\delta(w)$ is the Dirac delta,

$$\sigma_w^i = \frac{\sqrt{-2\hat{q}_0}}{2\hat{\Delta}\sigma_i} \tag{32}$$

is the (estimated) variance of the ith return,

$$w_i^+ = \frac{\lambda - \eta^+}{2\sigma_i^2\hat{\Delta}} \tag{33}$$

is the center of the Gaussian distribution of the (estimated) positive weight i,

$$w_i^- = \frac{\lambda + \eta^-}{2\sigma_i^2\hat{\Delta}} \tag{34}$$

is the same for negative weight i, and finally,

$$n_0 = \frac{1}{N}\sum_i \left[\Phi\left(\frac{\lambda+\eta^-}{\sigma_i\sqrt{-2\hat{q}_0}}\right) - \Phi\left(\frac{\lambda-\eta^+}{\sigma_i\sqrt{-2\hat{q}_0}}\right)\right] \tag{35}$$

is the density of the assets whose weights are set to zero by the regularizer.

We wish to make an important remark here: the right hand side of (19) is just $1 - n_0$. This will prove to be the key to the mapping between the regularized and unregularized cases.

Let us record the condensate density n_0 also for the special case when short positions are excluded ($\eta^- \to \infty$), but long positions are not penalized ($\eta^+ = 0$):

$$n_0 = \frac{1}{N} \sum_i \left[1 - \Phi\left(\frac{\lambda}{\sigma_i \sqrt{-2\hat{q}_0}} \right) \right] . \tag{36}$$

From (36), we can see that, since $\Phi(x)$ is monotonic increasing and, for $x \geq 0$, concave, the contribution to n_0 from assets with larger σ_is is larger than that from smaller σ_is. This means that in the no-short limit, the regularizer ℓ_1 eliminates more volatile assets with larger probability than the less volatile ones. Thus, we can think of the no-short constraint as a smooth upper cutoff in volatility. This is not true in the generic case (35), where the contributions of the small and large volatility items depend on the order parameters and the regularizer's slopes η^+ and η^- in a complicated manner: the probability of an asset with volatility σ_i to be removed is given by the difference of the two term in (35) under the sum. We do not wish to analyze this situation in detail, apart from the remark that a sufficiently large η^- generally favors the elimination of large volatility items.

The integral of $p(w)$ is, of course, 1. Its first moment, $\langle w^* \rangle_{\sigma,z}$, works out to be the same as (18):

$$\langle w^* \rangle_{\sigma,z} = 1 . \tag{37}$$

The second moment of the weight distribution is readily obtained as

$$\langle (w^*)^2 \rangle_{\sigma,z} = -\frac{\hat{q}_0}{\hat{\Delta}^2} \frac{1}{N} \sum_i \frac{1}{\sigma_i^2} \left[W\left(\frac{\lambda - \eta^+}{\sigma_i \sqrt{-2\hat{q}_0}} \right) + W\left(-\frac{\lambda + \eta^-}{\sigma_i \sqrt{-2\hat{q}_0}} \right) \right] . \tag{38}$$

The variance of the weight distribution is then

$$\langle (w^*)^2 \rangle_{\sigma,z} - \left(\langle w^* \rangle_{\sigma,z} \right)^2 , \tag{39}$$

which is equal to $q_0 - 1$, when the variances of the assets are all equal to 1. For a portfolio with different σ_i's, however, the relevant quantity that determines the out-of-sample estimate of ES is not the second moment of the weight distribution, but the true variance of the ith asset multiplied by the estimated portfolio weights squared and summed over the different assets, that is

$$\langle \sigma^2 (w^*)^2 \rangle_{\sigma,z} , \tag{40}$$

which is precisely q_0 as given in (20), and this is the quantity (multiplied by the correction as in (7)) that enters the formula for the out-of-sample estimate of ES in (8). For a not too inhomogeneous portfolio, the difference between the second moment of the weight distribution and q_0 is not significant, so we can think of q_0 as a measure of the variance of the portfolio.

Now we are ready to consider various special cases.

3.1. The Limit of Complete Information

When we have many observations (very long time series, $T \to \infty$) relative to the dimension N of the portfolio, we are in the $r = N/T \to 0$ limit. As we have already mentioned, this also corresponds to the "chemical potential" λ going to infinity. Obviously, in this limit, the regularizer plays no role.

We need the asymptotic behavior of the functions appearing in our stationary conditions: for $x \to \infty$, $\Phi(x) \to 1$, $\Psi(x) \sim x$, and $W(x) \sim x^2/2$, while for $x \to -\infty$, all three vanish exponentially.

Then from (18) we have

$$1 = \frac{\lambda}{2\hat{\Delta}} \frac{1}{N} \sum_i \frac{1}{\sigma_i^2} \,. \tag{41}$$

From (19)

$$2\Delta\hat{\Delta} = 1 \,. \tag{42}$$

Combining the two:

$$1 = \lambda\Delta \frac{1}{N} \sum_i \frac{1}{\sigma_i^2} \,. \tag{43}$$

We know from (1) and (28) that λ must be inversely proportional to r when $r \to 0$. It follows that $\Delta \sim r$ for small r.

Then, from (20) we find

$$q_0 = \Delta^2 \lambda^2 \frac{1}{N} \sum_i \frac{1}{\sigma_i^2} \,. \tag{44}$$

Combined with the previous equation, this gives

$$q_0 = \frac{1}{\frac{1}{N}\sum_i \frac{1}{\sigma_i^2}} \,. \tag{45}$$

The "true" ($r \to 0$) value of the order parameter q_0 is thus determined by the structural constant $\frac{1}{N}\sum_i \frac{1}{\sigma_i^2}$, which is given by the variances of the returns σ_i^2. This is in accord with the corresponding result found in the case of the ℓ_1-regularized variance risk measure [21,29]. The above result for q_0 also means that the quantity $\tilde{q}_0$ introduced in (7) is equal to 1, and according to (8) the out-of-sample estimate of ES is equal to its true value $\mathrm{ES}^{(0)}$, the estimation error is zero—an obvious result for the case of complete information.

From (23) with $\Delta \to 0$ we obtain $\alpha = \Phi(\epsilon/\sqrt{q_0})$, or

$$\epsilon = \Phi^{-1}(\alpha)\sqrt{q_0} \,. \tag{46}$$

Now from (21) we get $r = \Phi'\left(\frac{\epsilon}{\sqrt{q_0}}\right)\frac{\Delta}{\sqrt{q_0}}$, or

$$\Delta = r\sqrt{q_0}\frac{1}{\frac{1}{\sqrt{2\pi}}e^{-\epsilon^2/2q_0}} \,. \tag{47}$$

However, then we have found

$$\lambda = \frac{q_0}{\Delta} = \frac{1}{r}\frac{1}{\sqrt{2\pi}}e^{-\epsilon^2/2q_0}\sqrt{q_0} = \frac{1}{r}\frac{1}{\sqrt{2\pi}}e^{-\left(\Phi^{-1}(\alpha)\right)^2/2}\sqrt{q_0} \,. \tag{48}$$

Since $\lambda = f$ and $\mathrm{ES} = fr/(1-\alpha)$, we have the $r \to 0$ limit (the true value) of ES:

$$\mathrm{ES}^{(0)} = \frac{1}{1-\alpha}\frac{1}{\sqrt{2\pi}}e^{-\left(\Phi^{-1}(\alpha)\right)^2/2}\sqrt{q_0} \,. \tag{49}$$

We record the $r \to 0$ limits of the two auxiliary variables, $\hat{\Delta}$ and $\hat{q}$, for completeness:

$$\hat{\Delta} = \frac{1}{2r\sqrt{q_0}} \frac{1}{\sqrt{2\pi}} e^{-\epsilon^2/2q_0} \tag{50}$$

and

$$\hat{q}_0 \sim -\frac{1}{r}\,, \tag{51}$$

with a coefficient that will not be needed in the following.

Let us turn to the distribution of weights now.

In the $r \to 0$ limit, the widths of the Gaussians in (30) all vanish, so the Gaussians become delta functions:

$$p = \frac{1}{N}\sum_i \delta(w - w_i^+)\theta(w) + \frac{1}{N}\sum_i \delta(w - w_i^-)\theta(-w)\,. \tag{52}$$

In the $r \to 0$ limit, the weights are all positive, so the second sum disappears.

For the weights, w_i^+ we find

$$w_i^+ \simeq \frac{\lambda}{2\sigma_i^2\hat{\Delta}} = \frac{\lambda\Delta}{\sigma_i^2} = \frac{1}{\sigma_i^2}\frac{1}{\frac{1}{N}\sum_k \frac{1}{\sigma_k^2}}\,. \tag{53}$$

They sum to N, as stipulated.

The variance of a linear combination of independent random variables with averages w_i^+ and variances σ_i^2 is

$$\sigma_p^2 = \sum_i \left(w_i^+\right)^2 \sigma_i^2 = \frac{N}{\frac{1}{N}\sum_k \frac{1}{\sigma_k^2}}\,. \tag{54}$$

Now we recognize the meaning of the (true value of the) order parameter q_0: it is the normalized (to $\mathcal{O}(1)$) variance of the portfolio. This also explains the correction factor appearing in (7). We also see that (46) and (49) are the standard expressions for Value at Risk and Expected Shortfall indeed.

We emphasize again that all the results presented in this subsection are only valid in the $r \to 0$ limit when we are dealing with a finite dimension N and infinitely long time series T.

For finite r, the sample fluctuations start to broaden the delta spikes in the distribution of weights, the condensation of zero weights begins, λ decreases, and all the formulae above become considerably more complicated. We turn to this situation in the next subsections.

By now, we have learned everything that was to be learned from keeping the variances σ_i different, in particular the tendency of the elimination of the most volatile assets by the regularizer in the case of restriction of short selling. In order to simplify the presentation and avoid the appearance of very large and hardly transparent formulae, henceforth we set all the σ_i's equal to 1. We stress, however, that the main message of this paper, namely the existence of a mapping between the regularized and unregularized cases, depends only on the structure of the equations, and works also with different σ's.

3.2. Without Regularization

In this subsection, we set $\eta^+ = \eta^- = 0$, that is we consider our problem without regularization, and according to what has just been said, put $\sigma_i = 1$. We will make use of the identities

$$\Phi(x) + \Phi(-x) = 1 \tag{55}$$
$$\Psi(x) + \Psi(-x) = x \tag{56}$$
$$W(x) + W(-x) = \frac{1}{2}(x^2 + 1)\,. \tag{57}$$

The free energy (17) becomes

$$f = \lambda - \frac{\alpha\epsilon}{r} - \Delta\hat{q}_0 - \hat{\Delta}q_0 - \frac{\Delta}{2r} + \frac{q_0}{r\Delta}\left[W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right] - \frac{\lambda^2}{4\hat{\Delta}} + \frac{\hat{q}_0}{2\hat{\Delta}}. \tag{58}$$

For the saddle point equations, we find:

$$1 = \frac{\lambda}{2\hat{\Delta}}\,, \tag{59}$$

$$2\Delta\hat{\Delta} = 1\,, \tag{60}$$

$$q_0 = \frac{\lambda^2}{4\hat{\Delta}^2} - \frac{\hat{q}_0}{2\hat{\Delta}^2}\,, \tag{61}$$

$$2r\Delta\hat{\Delta} = r = \Phi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - \Phi\left(\frac{\epsilon}{\sqrt{q_0}}\right), \tag{62}$$

$$\alpha = \frac{\sqrt{q_0}}{\Delta}\left[\Psi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - \Psi\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right], \tag{63}$$

$$\hat{q}_0 + \frac{2q_0\hat{\Delta}}{\Delta} + \frac{\alpha\epsilon}{r\Delta} + \frac{1}{2r} - \frac{q_0}{r\Delta^2}\left[W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right] = 0\,. \tag{64}$$

These equations are rather similar to their counterparts in the previous subsection, but of course $r \to 0$ is not assumed here. As for their solutions, they were discussed and illustrated in several figures in [12], therefore we will not dwell upon them here. (Some results will be given in Section 3.6.) Instead, we write up the corresponding equations in the case where no short positions are allowed and make a term-by-term comparison between the two sets of equations.

3.3. No Short Selling

Short positions will be excluded by imposing infinite penalty on them by letting η^- go to infinity. The functions $\Phi(x)$, $\Psi(x)$, and $W(x)$ all vanish when $x \to -\infty$. Long positions will not be penalized, so we set $\eta^+ = 0$.

The free energy becomes

$$f = \lambda - \frac{\alpha\epsilon}{r} - \Delta\hat{q}_0 - \hat{\Delta}q_0 - \frac{\Delta}{2r} + \frac{q_0}{r\Delta}\left[W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right] \tag{65}$$

$$+ \frac{\hat{q}_0}{\hat{\Delta}}W\left(\frac{\lambda}{\sqrt{-2\hat{q}_0}}\right). \tag{66}$$

The stationary conditions now read as:

$$1 = \frac{\sqrt{-2\hat{q}_0}}{2\hat{\Delta}}\Psi\left(\frac{\lambda}{\sqrt{-2\hat{q}_0}}\right), \tag{67}$$

$$2\Delta\hat{\Delta} = \Phi\left(\frac{\lambda}{\sqrt{-2\hat{q}_0}}\right), \tag{68}$$

$$q_0 = -\frac{\hat{q}_0}{\hat{\Delta}^2}W\left(\frac{\lambda}{\sqrt{-2\hat{q}_0}}\right), \tag{69}$$

$$2r\Delta\hat{\Delta} = \Phi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - \Phi\left(\frac{\epsilon}{\sqrt{q_0}}\right), \tag{70}$$

$$\alpha = \frac{\sqrt{q_0}}{\Delta}\left[\Psi\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - \Psi\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right], \tag{71}$$

$$r\left(\hat{q}_0 + \frac{2q_0\hat{\Delta}}{\Delta}\right) + \frac{\alpha\epsilon}{\Delta} + \frac{1}{2} - \frac{q_0}{\Delta^2}\left[W\left(\frac{\Delta+\epsilon}{\sqrt{q_0}}\right) - W\left(\frac{\epsilon}{\sqrt{q_0}}\right)\right] = 0\,, \tag{72}$$

the last equation being the same as (64), just multiplied by r.

In the distribution of weights in (30), the second sum of Gaussians will disappear, because for $\eta^- \to \infty$, all the weights (34) go to infinity. The weights (33) become

$$w_i^+ = \frac{\lambda}{2\hat{\Delta}}\,, \tag{73}$$

while the density of zero weights is now

$$n_0 = 1 - \Phi\left(\frac{\lambda}{\sqrt{-2\hat{q}_0}}\right), \tag{74}$$

which with (68) leads to

$$1 - n_0 = 2\Delta\hat{\Delta}\,. \tag{75}$$

From (74), we see that $n_0 = 0$ for $r = 0$ and increases as λ decreases, until it reaches its maximal value $1/2$ when λ vanishes. Mathematically, there is nothing to prevent us from continuing to increase r and driving λ to negative values, which would allow n_0 to grow beyond $1/2$, up to $n_0 = 1$, but a negative λ would cause the free energy and thus also ES to change sign—an extreme case of "in-sample optimism", entirely due to the lack of sufficient information. We consider such a situation "unphysical", and never go beyond the point where λ (or $\lambda - \eta^+$ if $\eta^+ > 0$) vanishes anywhere in this paper.

3.4. No-Short Mapping

We are now ready to spell out the mapping between the no-short case and the unregularized one.

The first point to notice is that the only difference between Equation (62) valid in the unregularized case and its counterpart (70) in the no-short case (combined with (75)) appears on their left hand side: the terms r and $(1 - n_0)r$, respectively. This suggests to introduce an effective r:

$$r_{\text{eff}} = (1 - n_0)r\,. \tag{76}$$

Now $r = N/T$, and n_0 is the density of the assets removed by the regularizer, thus $(1-n_0)r = \frac{N-N_0}{T}$ is the number of surviving assets divided by the length of the time series. As r_{eff} increases from zero to $1/2$, r will increase between zero and 1.

Inspired by the connection between r and r_{eff}, we compare the two sets of equations and recognize that, in fact, the whole system of saddle point equations can be mapped from the regularized case to the unregularized one. A variable that appears in all the subsequent equations is

$$z = \frac{\lambda}{\sqrt{-2\hat{q}_0}}\,, \tag{77}$$

where the variables λ and $\hat{q}_0$ are those that appear in the no-short equations.

Then the connection between the order parameters belonging to the two cases is the following:

$$q_0 = q_0^{\text{eff}} \frac{z}{\Psi(z)}\,, \tag{78}$$

$$\Delta = \Delta_{\text{eff}} \sqrt{\frac{z}{\Psi(z)}}\,, \tag{79}$$

$$\epsilon = \epsilon_{\text{eff}} \sqrt{\frac{z}{\Psi(z)}}\,, \tag{80}$$

$$\lambda = \lambda_{\text{eff}} \sqrt{\frac{z}{\Psi(z)}}\Phi(z)\,, \tag{81}$$

$$\hat{q}_0 = \hat{q}_0^{\text{eff}}\Phi(z)\,, \tag{82}$$

$$\hat{\Delta} = \hat{\Delta}_{\text{eff}} \sqrt{\frac{\Psi(z)}{z}}\Phi(z)\,. \tag{83}$$

A direct substitution shows that if the order parameters on the left hand sides of the above equations satisfy the no-short equations, then the effective variables satisfy the unregularized ones, provided we also replace r with r_{eff}. In particular, the contour maps of the unregularized order parameters presented in [12] can be taken over and simply blown up by a factor $\frac{1}{1-n_0}$ to obtain the contour maps of the no-short variables. Given the relation between q_0 and the estimation error, we see that the mapping also means that a given error belongs to a larger r in the no-short case than in the unregularized one, in other words, the no-short constrained problem demands $(1-n_0)$ times less data (shorter time series) than the unregularized one.

One may wonder whether this mapping expresses some symmetry of the problem, that is whether the free energy functional is invariant under this mapping. The answer is no: the mapping works only in the saddle point equations, it is a property of the stationary point.

It is important to learn the range of this transformation. In the limit $r \to 0$, the transformation is the identity, but this is trivial: when we have complete information, the regularizer does not play any role. It is more interesting to consider the vicinity of the phase transition in the unregularized case, where q_0^{eff} and Δ_{eff} diverge. These divergences are removed by the mapping, no singularity is found in the no-short case. This is in accord with [18]: the infinite penalty on short positions precludes the phase transition and no singularity shows up in q_0 , Δ, or ϵ. Mathematically, we can continue the unregularized solutions into the non-feasible region beyond the phase boundary, but they make no sense there (for example, q_0 changes sign, Δ and ϵ become imaginary, etc.), while their mapped counterparts continue to behave reasonably. According to (76), when r_{eff} reaches the critical point $r_c(\alpha)$, the corresponding value of r in the no-short problem will be twice as large, so the whole phase diagram is multiplied by a

factor 2. Beyond the mapped phase boundary the regularized solutions still survive, but their meaning becomes questionable, because the free energy, hence also ES change sign. As noted in the previous Subsection, we refrain from the discussion of this unphysical region.

3.5. Mapping for Generic ℓ_1 Constraint

The mapping between the generic ℓ_1-constrained ES optimization and the unregularized one is a straightforward generalization of the results in the previous Subsection. The mapping is made more complicated because of the sums and differences of the Ψ, Φ, and W functions appearing on the right hand side of Equations (18)–(20). We introduce the following notation for these combinations:

$$A_\Psi = \Psi\left(\frac{\lambda - \eta^+}{\sqrt{-2\hat{q}_0}}\right) - \Psi\left(-\frac{\lambda + \eta^-}{\sqrt{-2\hat{q}_0}}\right), \tag{84}$$

$$A_\Phi = \Phi\left(\frac{\lambda - \eta^+}{\sqrt{-2\hat{q}_0}}\right) + \Phi\left(-\frac{\lambda + \eta^-}{\sqrt{-2\hat{q}_0}}\right), \tag{85}$$

and

$$A_W = W\left(\frac{\lambda - \eta^+}{\sqrt{-2\hat{q}_0}}\right) + W\left(-\frac{\lambda + \eta^-}{\sqrt{-2\hat{q}_0}}\right), \tag{86}$$

where we have set all the $\sigma_i = 1$.

In terms of these quantities the generic map reads as

$$q_0 = q_0^{\text{eff}} \frac{2A_W - A_\Phi}{(A_\Psi)^2}\,, \tag{87}$$

$$\Delta = \Delta_{\text{eff}} \frac{\sqrt{2A_W - A_\Phi}}{A_\Psi}\,, \tag{88}$$

$$\epsilon = \epsilon_{\text{eff}} \frac{\sqrt{2A_W - A_\Phi}}{A_\Psi}\,, \tag{89}$$

$$\lambda = \lambda_{\text{eff}} \frac{zA_\Phi}{\sqrt{2A_W - A_\Phi}}\,, \tag{90}$$

$$\hat{q}_0 = \hat{q}_0^{\text{eff}} A_\Phi\,, \tag{91}$$

$$\hat{\Delta} = \hat{\Delta}_{\text{eff}} \frac{A_\Phi A_\Psi}{\sqrt{2A_W - A_\Phi}}\,. \tag{92}$$

For the condensate density n_0, we have

$$1 - n_0 = A_\Phi\,, \tag{93}$$

and for the effective aspect ratio

$$r_{\text{eff}} = 2r\Delta\hat{\Delta} = rA_\Phi = (1 - n_0)r\,. \tag{94}$$

As before, if the order parameters satisfy the regularized stationarity conditions (18)–(27) (with $\sigma_i = 1$), then the effective parameters will satisfy the unregularized Equations (59)–(64), and vice versa.

Note that the above equations remain invariant if we redefine λ as $\lambda - \eta^+$ and η^- as $\eta^- + \eta^+$. So we can set $\eta^+ = 0$ and $\eta^- + \eta^+ = \eta$ without loss of generality. We will use this setup in the following, in order to reduce the number of parameters when solving the stationarity equations.

3.6. Solutions for the Order Parameters

Except for a few exceptional points, it is impossible to obtain the solutions of the stationarity equations in closed, analytical form, but it is perfectly possible to get them numerically, by a computer. (The case of $\alpha = 1$ is exceptional in several respects and will not be considered here.) In the following, the solutions will be presented in graphical form.

Figure 1 exhibits three special lines, belonging to three different cases: the unregularized case, the one with a finite regularizer, and the one with a no-short constraint.

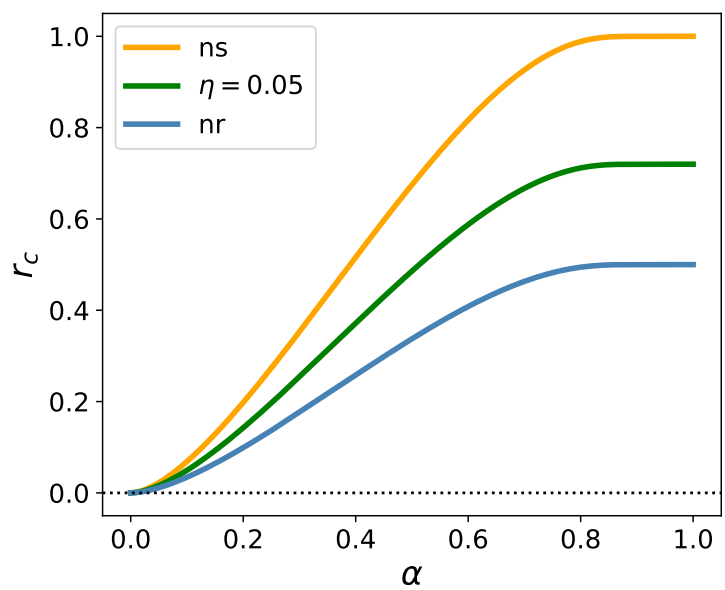

Figure 1. The boundary of the region where the optimization of ES is feasible in the unregularized case (nr); its image under the map for a finite $\eta^- = 0.05$, $\eta+ = 0$ regularizer; and the same under the no-short map (ns).

The blue line is the upper boundary of the region where the optimization of unregularized ES is feasible. This line was first determined in [10]. It is a phase boundary, along which a phase transition takes place: q_0, Δ, and ϵ diverge here, while λ becomes zero. The unregularized equations can be solved also above this line, up to the horizontal line at $r = 1$ (not shown in the Figure), but the solutions are meaningless: q_0 is negative, while λ, Δ, and ϵ become imaginary. The unregularized equations do not have any solution above $r = 1$.

The green line is the image of the unregularized phase boundary under the mapping described in the previous Subsection, and corresponds to a one-sided regularizer with $\eta^- = 0.05$, $\eta+ = 0$. There is no phase transition when we cross this line, the order parameters remain smooth, finite quantities, but λ (along with the free energy and the in-sample estimate of ES) changes sign, rendering the solution in the region above the green line "unphysical". Nevertheless, if we keep following the solutions beyond the green line we can go up to the image of the $r = 1$ line (mapped into $r \to \infty$), where q_0 and Δ will ultimately diverge. The region between the green line and the image of the $r = 1$ line has an intricate structure, but because it corresponds to negative risk, it is of no interest for us in the present context.

In the no-short case, there is always a solution with the order parameters remaining finite all the way up to infinity, which is the image of the $r = 1$ line under the no-short map. However, as we cross the orange line, λ changes sign, and the region beyond it is meaningless again. The orange line is the unregularized phase boundary (blue line) blown up by a factor $\frac{1}{1-n_0} = 2$. All this is in accord with the picture described in [18] in that the no-short constraint eliminates the critical line. The solutions becoming unphysical beyond a certain r-range could not be foreseen on the basis of the analysis in [18].

Figure 2 shows the η-dependence of q_0 and the density of the zero weights n_0 at criticality, and that of the value of the critical r. In the unregularized case ($\eta \to 0$), $q_0 \to \infty$, while in the no-short case ($\eta \to \infty$) $q_0 \to \pi$. At $\alpha = 0.975$, the value of the critical r_c increases from $r_c \approx 1/2$ in the unregularized case to $\approx$1 for the no-short case. The proportion of the assets eliminated from the portfolio (the condensate density) goes from zero for $\eta = 0$ to $1/2$ for large η.

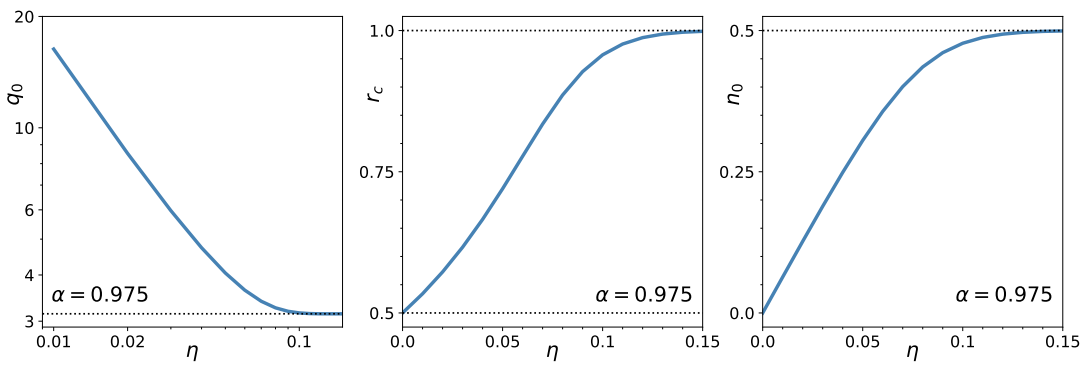

Figure 2. Dependence of q_0 at r_c (**left**), critical point (**middle**), and proportion of zero weights at r_c (**right**) as a function of the regularization strength, $\eta^- = \eta$ ($\eta^+ = 0$). Note the logarithmic scale in the left panel.

In Figure 3, we display the r-dependence of q_0, Δ, and λ for the three cases: unregularized, regularized, and no-short. Without regularization, q_0 and Δ increase with r and diverge at an r_c slightly less than $\frac{1}{2}$; while λ decreases from infinity at $r = 0$ to zero at r_c. (The confidence limit α is set at its regulatory value 0.975 in these figures.) Under the regularizer $\eta^- = 0.05$, $\eta^+ = 0$, q_0, and Δ increases up to the r where λ vanishes. The situation is similar for an infinitely strong (no-short) regularizer, with the limiting value of $q_0 = \pi$ and $\lambda = 0$ at $r \approx 1$.

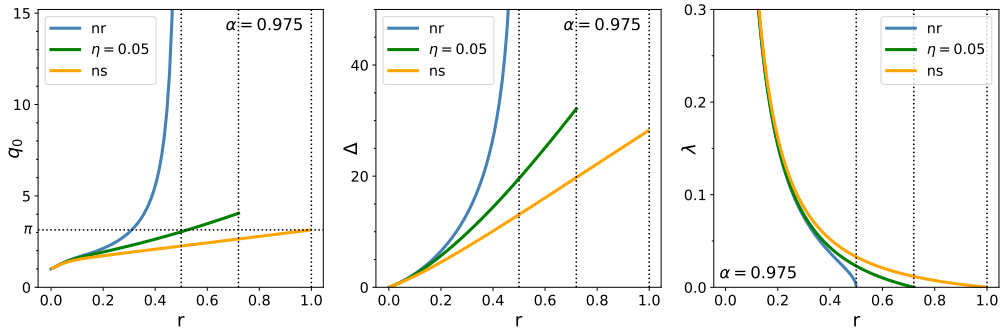

Figure 3. Dependence of q_0(**left**), Δ (**middle**) and "chemical potential" λ (**right**) on $r = N/T$, for the unregularized (blue), $\eta^- = 0.05, \eta^+ = 0$ regularized (green), and no-short (yellow) cases.

The left panel in Figure 4 shows the relative out-of-sample estimation error, which is related to the out-of-sample estimate of ES by (8) ($\tilde{q}_0 = q_0$ now, as we have set all the σ_i =1).

These curves are similar to the curves of q_0 in the previous Figure. It can be seen that the curves of the relative estimation error run very close to each other for small values of r: there is no substantial reduction of the error in this range. Where they fan out and the effect of regularization starts to be felt (say around $r = 0.1$), the relative error is already about 20%.

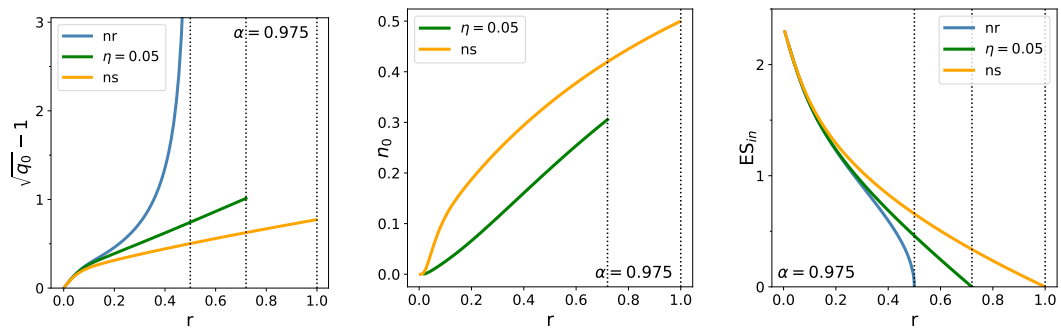

Figure 4. Dependence of the out-of-sample estimation error (**left**), proportion of zero weights (**center**), and in-sample ES (**right**) on $r = N/T$, for the non-regularized (blue), $\eta^- = \eta$ ($\eta^+ = 0$) regularized (green), and no-short (orange) cases.

The middle panel in Figure 4 shows the behavior of the density of zero weights as function of r for the finite η-regularized and the no-short cases. In the no-short case, n_0 reaches its maximal value $\frac{1}{2}$ at $r \approx 1$ (for $\alpha = 0.975$) where λ vanishes. For a regularizer of finite strength, it always remains below $\frac{1}{2}$.

The right panel in Figure 4 displays the behavior of the in-sample estimate of ES for the three cases. This quantity is directly related to λ through (1) and (28). The monotonic and fast decay of these curves demonstrates what is called in-sample optimism, a strong underestimation of risk.

4. Discussion

In the preceding Section we compared the behavior of the order parameters in the three instances considered in this paper: the case of the unregularized, the ℓ_1-regularized, and the no-short constrained Expected Shortfall optimization. We have seen that without regularization, there is a phase transition as we cross the phase boundary $r_c(\alpha)$ shown in Figure 1 with Δ, q_0, and ϵ diverging here, as known since the paper [10]. In contrast, the infinite penalty on short positions suppresses this phase transition, while an ℓ_1 regularizer with finite slopes only shifts the phase boundary. These facts were also known from earlier work [14,18]. However, the picture has turned out to be more complicated than envisaged in [18]. The numerical solution for the order parameters performed in this paper has revealed that new characteristic lines emerge both in the case of finite regularization and the no-short constraint, along which the order parameter λ and, consequently, the free energy and the in-sample estimate of Expected Shortfall change sign. We have determined the position of these new characteristic lines: in the no-short case the new line is the curve $2r_c(\alpha)$, for a finite regularizer it is $\frac{r_c(\alpha)}{1-n_0}$, where $n_0 \leq \frac{1}{2}$. We have omitted the detailed analysis of the regions above these lines, where the estimated risk becomes negative. Instead, we confined ourselves to merely pointing out that the critical line for the no-short constraint is projected out to infinity, so the phase transition is removed indeed, while for a finite slope regularizer the critical line is shifted into the unphysical, negative risk region, where for some values of the regularizer's strength η, it even develops two branches.

We have also found the behavior of the various order parameters, most notably that of q_0 that determines the out-of-sample estimation error of ES, the free energy that gives the

in-sample estimator, and the susceptibility-like quantity Δ, and displayed their behavior for the three cases studied here. It is satisfactory to see that q_0 and Δ remain finite up to the new characteristic lines, that is, the regularizer acts as expected: it suppresses the divergent sample fluctuations in the optimization of ES. Unfortunately, this suppression is not strong enough to bring down the estimation error to acceptable values, except for the range of small $r = \frac{N}{T}$ ratios where it demands far too long time series for any realistic N, and where r is small already without any regularization.

What is the meaning of this phase transition? As analyzed in [8,26] it follows from the coherence axioms that coherent risk measures, including ES, are unstable in the sense that whenever an asset or a combination of assets in the portfolio stochastically dominates the others in a given sample, the investor can take an extremely large long position in the dominant asset and compensate this with an appropriately large short position, without violating the budget constraint. This means that the weight of the dominant asset runs away practically to infinity, resulting in an arbitrarily large negative value of the risk measure. This is a mirage of an arbitrage, which can disappear in the next sample, or change into another arbitrage with a different weight running away to infinity. In practice, there are always constraints that prevent such a divergence from taking place. The ban on short selling is just this sort of constraint. The runaway solutions try to escape, but get arrested at the walls constituted by the constraint, in the case of a no-short ban, at the coordinate planes. This is how the condensate of zero weights builds up. This mechanism is the stronger the larger the ratio $r = N/T$.

There is nothing surprising about solutions sitting on the constraint-walls or at corners in a linearly programmable problem, such as the optimization of ES. In the usual applications of linear programming, the constraints typically express some physical limitation like a finite amount of resources, material or labor, etc. In the present finance problem, such a finite resource would be the limited budget, but if short selling is not constrained, the budget in itself cannot prevent runaway solutions. The ban on short positions corresponds to an infinitely strong ℓ_1 regularizer, which, combined with the budget constraint, is already sufficient to take care of the runaway solutions. So, with a no-short ban on, we can increase r (that is the dimension, or decrease the amount of data) without any mathematical contradiction showing up; neither q_0 nor Δ will diverge. It is clear, however, that the solution based on less and less information becomes increasingly meaningless. In these circumstances, the optimization will not tell us anything useful about the structure of the market, it will be determined more and more by the constraint.

What we regard as the most intriguing result of this paper is the existence of a mapping between the regularized and the unregularized problems.

Author Contributions: Conceptualization, G.P., I.K. and F.C.; formal analysis, G.P. and I.K.; funding acquisition, F.C.; investigation, G.P., I.K. and F.C.; writing—original draft preparation, I.K.; writing—review and editing, G.P. and F.C.; visualization, G.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We are indebted to Susanne Still and Matteo Marsili for collaboration and useful discussions years ago on joint works preceding the present one. Although they did not participate in this work, their ideas have remained a source of inspiration for us. I.K. is obliged to Risi Kondor for several enlightening discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Markowitz, H. Portfolio selection. *J. Financ.* **1952**, *7*, 77–91.
2. Morgan, J. *Riskmetrics Technical Manual*; JP Morgan: New York, NY, USA, 1995.
3. Basel Committee on Banking Supervision. *Overview of the Amendment to the Capital Accord to Incorporate Market Risks*; Bank for International Settlements: Basel, Switzerland, 1996.
4. Artzner, P.; Delbaen, F.; Eber, J.M.; Heath, D. Coherent Measures of Risk. *Math. Financ.* **1999**, *9*, 203–228. [CrossRef]
5. Acerbi, C.; Tasche, D. Expected Shortfall: A Natural Coherent Alternative to Value at Risk. *Econ. Notes* **2002**, *31*, 379–388. [CrossRef]
6. Pflug, G.C. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization*; Uryasev, S., Ed.; Springer: Boston, MA, USA, 2000; pp. 272–281.
7. Basel Committee on Banking Supervision. Minimum Capital Requirements for Market Risk. 2016. Available online: https://www.bis.org/bcbs/publ/d352.htm (accessed on 23 April 2021).
8. Kondor, I.; Varga-Haszonits, I. Instability of portfolio optimization under coherent risk measures. *Adv. Complex Syst.* **2010**, *13*, 425–437. [CrossRef]
9. Kondor, I.; Pafka, S.; Nagy, G. Noise sensitivity of portfolio selection under various risk measures. *J. Bank. Financ.* **2007**, *31*, 1545–1573. [CrossRef]
10. Ciliberti, S.; Kondor, I.; Mézard, M. On the Feasibility of Portfolio Optimization under Expected Shortfall. *Quant. Financ.* **2007**, *7*, 389–396. [CrossRef]
11. Kondor, I.; Caccioli, F.; Papp, G.; Marsili, M. Contour Map of Estimation Error for Expected Shortfall. 2015. Available online: http://ssrn.com/abstract=2567876 and http://arxiv.org/abs/1502.0621 (accessed on 23 April 2021).
12. Caccioli, F.; Kondor, I.; Papp, G. Portfolio optimization under expected shortfall: Contour maps of estimation error. *Quant. Financ.* **2018**, *18*, 1295–1313. [CrossRef]
13. Bühlmann, P.; Van De Geer, S. *Statistics for High-Dimensional Data: Methods, Theory and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011.
14. Caccioli, F.; Still, S.; Marsili, M.; Kondor, I. Optimal liquidation strategies regularize portfolio selection. *Eur. J. Financ.* **2013**, *19*, 554–571. [CrossRef]
15. Papp, G.; Caccioli, F.; Kondor, I. Variance-bias trade-off in portfolio optimization under Expected Shortfall with ℓ_2 regularization. *J. Stat. Mech. Theory Exp.* **2019**, *2019*, 013402. [CrossRef]
16. Still, S.; Kondor, I. Regularizing portfolio optimization. *New J. Phys.* **2010**, *12*, 075034. [CrossRef]
17. Rockafellar, R.T.; Uryasev, S. Optimization of Conditional Value-at-Risk. *J. Risk* **2000**, *2*, 21–41. [CrossRef]
18. Caccioli, F.; Kondor, I.; Marsili, M.; Still, S. Liquidity Risk and Instabilities In Portfolio Optimization. *Int. J. Theor. Appl. Financ.* **2016**, *19*, 1650035. [CrossRef]
19. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, 2nd ed.; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 2008.
20. Mézard, M.; Parisi, G.; Virasoro, M.A. *Spin Glass Theory and Beyond*; World Scientific Lecture Notes in Physics Volume 9; World Scientific: Singapore, 1987.
21. Kondor, I.; Papp, G.; Caccioli, F. Analytic approach to variance optimization under an ℓ_1 constraint. *Eur. Phys. J.* **2019**, *92*, 8. [CrossRef]
22. Kempf, A.; Memmel, C. Estimating the global minimum variance portfolio. *Schmalenbach Bus. Rev.* **2006**, *58*, 332–348. [CrossRef]
23. Basak, G.K.; Jagannathan, R.; Ma, T. A jackknife estimator for tracking error variance of optimal portfolios constructed using estimated inputs. *Manag. Sci.* **2009**, *55*, 990–1002. [CrossRef]
24. Frahm, G.; Memmel, C. Dominating estimators for minimum-variance portfolios. *J. Econom.* **2010**, *159*, 289–302. [CrossRef]
25. Ciliberti, S.; Mézard, M. Risk minimization through portfolio replication. *Eur. Phys. J.* **2007**, *B 57*, 175–180. [CrossRef]
26. Varga-Haszonits, I.; Kondor, I. The instability of downside risk measures. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P12007. [CrossRef]
27. Shinzato, T. Minimal investment risk of portfolio optimization problem with budget and investment concentration constraints. *J. Stat. Mech. Theory Exp.* **2017**, *2017*, 023301. [CrossRef]
28. Kondor, I.; Papp, G.; Caccioli, F. Analytic solution to variance optimization with no short positions. *J. Stat. Mech. Theory Exp.* **2017**, *2017*, 123402. Available online: https://iopscience.iop.org/article/10.1088/1742-5468/aa9684 (accessed on 23 April 2021). [CrossRef]
29. Varga-Haszonits, I.; Caccioli, F.; Kondor, I. Replica approach to mean-variance portfolio optimization. *J. Stat. Mech. Theory Exp.* **2016**, *2016*, 123404. [CrossRef]