Notes on Alpha Stream Optimization

Zura Kakushadze[§†1]

[§] *Quantigic® Solutions LLC*
1127 High Ridge Road #135, Stamford, CT 06905 [2]
[†] *Free University of Tbilisi, Business School & School of Physics*
240, David Agmashenebeli Alley, Tbilisi, 0159, Georgia

(June 4, 2014; revised: March 18, 2015)

Abstract

In these notes we discuss investment allocation to multiple alpha streams traded on the same execution platform, including when trades are crossed internally resulting in turnover reduction. We discuss approaches to alpha weight optimization where one maximizes P&L subject to bounds on volatility (or Sharpe ratio). The presence of negative alpha weights, which are allowed when alpha streams are traded on the same execution platform, complicates the optimization problem. By using factor model approach to alpha covariance matrix, the original optimization problem can be viewed as a 1-dimensional root searching problem plus an optimization problem that requires a finite number of iterations. We discuss this approach without costs and with linear costs, and also with nonlinear costs in a certain approximation, which makes the allocation problem tractable without forgoing nonlinear portfolio capacity bound effects.

---

[1] Zura Kakushadze, Ph.D., is the President of Quantigic® Solutions LLC, and a Full Professor at Free University of Tbilisi. Email: zura@quantigic.com
[2] DISCLAIMER: This address is used by the corresponding author for no purpose other than to indicate his professional affiliation as is customary in publications. In particular, the contents of this paper are not intended as an investment, legal, tax or any other such advice, and in no way represent views of Quantigic Solutions LLC, the website www.quantigic.com or any of their other affiliates.

1 Introduction and Summary

Nowadays, technology allows to combine a large number of hedge fund alpha streams on the same trading platform.[3] Not only does this yield diversification, but also reduces transaction costs by crossing trades between different alpha streams.[4] Given a set of alpha streams, one has to decide how to allocate funds into these alphas, *i.e.*, the weights with which to combine them.[5] This is an old dilemma between greed and prudence – does one maximize the P&L, does one minimize the volatility, does one do something in between, or does one simply split the difference and maximize the Sharpe ratio instead? It all depends on one's risk tolerance. For instance, one could say, I want to maximize my P&L, but I also want my volatility to be capped, or the Sharpe ratio to be bounded from below, *etc.*

The purpose of these notes is to discuss aspects of optimizing weights for alpha streams. When we have a large number N of alphas α_i, this optimization is an N-dimensional problem and does not always have a simple solution. One way to tackle it is to try to essentially reduce it to a 1-dimensional problem, by considering simpler problems of maximizing P&L for fixed volatility or Sharpe ratio, or by maximizing (minimizing) Sharpe ratio (volatility) for fixed P&L, and then to try to find a desirable configuration using standard search techniques. This works in some cases, but not always. One issue is that, when alpha streams are combined on the same trading platform, some weights can be negative – this is to be contrasted with the case when alpha streams are traded on their individual trading platforms, *i.e.*, a fund of funds type of a case, where the weights are non-negative. The weights being signed quantities complicates things because the weight normalization condition reads

$$\sum_{i=1}^{N} |w_i| = 1 \qquad (1)$$

It is the modulus in this condition that is the source of the headache. Furthermore, when trades between different alphas are crossed, the resulting portfolio turnover reduces (compared with the case with no internal crossing) adding further complexity to the problem. In [57] we discussed alpha stream optimization via Sharpe ratio maximization in the presence of linear costs. There, because the Sharpe ratio is invariant[6] under the rescalings $w_i \to \lambda w_i$, the condition (1) does not affect anything, because it is "trivially" satisfied by finding a solution without such a constraint and then simply rescaling all w_i so they satisfy (1). In the case where one does not maximize the Sharpe ratio but employs a more complex optimization criterion, scale invariance is lost and (1) causes complications in the minimization of the objective function similar to those discussed in [57], but even trickier.

---

[3] For a partial list of hedge fund literature, see, *e.g.*, [1]-[20] and references therein.
[4] For a recent discussion, see [21].
[5] For a partial list of portfolio optimization and related literature, see, *e.g.*, [22]-[56] and references therein.
[6] More precisely this is the case with no costs or linear costs only.

One of our main observations is that, if the alpha covariance matrix is of a factor model form, then an optimum can be found via a finite iterative procedure for a practically interesting range of the P&L and Sharpe ratio/volatility.[7] In Section 2 we set up our notations. In Section 3 we discuss the case of a diagonal covariance matrix, where the aforesaid modulus issue is absent. We discuss a relaxation algorithm for finding quasi-optimal weights (which are generally expected to be close to the optimal solution when the number of alphas is large) in Subsection 3.1. In Subsection 3.2 we discuss an algorithm for finding the weights when the P&L is maximized while the Sharpe ratio is fixed. In Subsection 3.3 we discuss a solution minimizing volatility (maximizing the Sharpe ratio) with the P&L fixed. In Section 4 we discuss the case of a non-diagonal covariance matrix and a trick where one effectively diagonalizes it by utilizing synthetic tradable portfolios. The problem then reduces to an iterative procedure where one needs to solve for the effective investment level. This procedure, however, generally is expected to have convergence issues which are obscured by the diagonalization trick. In Section 5 we therefore discuss how to tackle the problem directly in the original w_i basis, without diagonalizing the covariance matrix, and discuss the problem of minimizing the volatility with the P&L fixed, which is solvable via a finite iterative procedure assuming a multi-factor form for the covariance matrix. More precisely, this is the case for the practically interesting range of the P&L above its value corresponding to the maximum of the Sharpe ratio. We then generalize this approach to include linear costs in Section 6 and also nonlinear costs (impact) in Section 7. In the latter case we employ the approximation discussed in [57]. Section 8 concludes the paper with some observations and comments.

2   Definitions

We have N alphas α_i, $i = 1, \ldots, N$. Each alpha is actually a time series $\alpha_i(t_s)$, $s = 0, 1, \ldots, M$, where t_0 is the most recent time. Below α_i refers to $\alpha_i(t_0)$.

Let C_{ij} be the covariance matrix of the N time series $\alpha_i(t_s)$. Let Ψ_{ij} be the corresponding correlation matrix, *i.e.*,

$$C_{ij} = \sigma_i \ \sigma_j \ \Psi_{ij} \tag{2}$$

where $\Psi_{ii} = 1$.

Alphas α_i are combined with weights w_i such that

$$\sum_{i=1}^{N} |w_i| = 1 \tag{3}$$

---

[7] More precisely, this is the case with no costs or linear costs, and also in the presence of nonlinear costs in a certain approximation.

Here the modulus accounts for the fact that, when alphas are combined on the same trading platform, some weights can be negative if the covariance matrix C_{ij} is not diagonal, even if all alphas are positive.

To begin with, we will ignore trading costs. Portfolio P&L, volatility and Sharpe ratio are given by

$$P = I \sum_{i=1}^{N} \alpha_i \ w_i \tag{4}$$

$$R = I \ \sqrt{\sum_{i,j=1}^{N} C_{ij} \ w_i \ w_j} \tag{5}$$

$$S = \frac{P}{R} \tag{6}$$

where I is the investment level. Our goal is to find the set of w_i for which[8]

$$P \to \max \tag{7}$$

$$S \geq S_{\min} \tag{8}$$

for a given Sharpe ratio lower bound $S_{\min}$.

In practice, we can approach this problem in the following ways. First, we can start with the maximal possible P&L – which means that all weights, other than the weight corresponding to the largest alpha, are zero – and find a relaxation by gradually reducing the P&L until (8) is satisfied. Second, we can solve a simplified problem $P \to \max$ for a fixed value of $S = S_*$, and then use a standard searching algorithm based on a discrete set of values of S_* within the desired precision. Third, we can solve another simplified problem $S \to \max$ for a fixed value of P, and then use a standard searching algorithm based on a discrete set of values of P within the desired precision. *I.e.*, the basic idea is to reduce the large N-dimensional problem to a 1-dimensional searching algorithm. In some cases this approach does work.

3 Diagonal Case

When C_{ij} is not diagonal, some weights can be negative even if all alphas are positive. The modulus in (3) complicates things. However, if C_{ij} is diagonal, $C_{ij} = \sigma_i^2 \ \delta_{ij}$, then, assuming all $\alpha_i \geq 0$ (which we can do without loss of generality), all $w_i \geq 0$ as well.[9] Our optimization problem simplifies as follows. Let

$$\widehat{\alpha}_i \equiv \nu_i \ \alpha_i \tag{9}$$

$$\widehat{w}_i \equiv \sigma_i \ w_i \tag{10}$$

$$\nu_i \equiv 1/\sigma_i \tag{11}$$

---

[8] In practice, one may have to restate this criterion – see Section 8.

[9] This is because a configuration with a negative w_ℓ has the same volatility as and a lower P&L than the configuration with the sign of w_ℓ flipped.

Then we have

$$P = I \sum_{i=1}^{N} \widehat{\alpha}_i \ \widehat{w}_i \tag{12}$$

$$R = I \sqrt{\sum_{i=1}^{N} \widehat{w}_i^2} \tag{13}$$

and the optimization problem now is (7) and (8) subject to

$$\sum_{i=1}^{N} \nu_i \ \widehat{w}_i = 1 \tag{14}$$

$$\widehat{w}_i \geq 0 \tag{15}$$

with all $\widehat{\alpha}_i \geq 0$. Below we discuss a simple relaxation algorithm for maximizing P&L for $S = S_*$.

3.1 Quasi-Optimal Weights

We can obtain a quasi-optimal solution as follows. Recall that all $\widehat{\alpha}_i$ are non-negative. First, let us label alphas such that $\alpha_1 = \max(\alpha_i, i = 1, \ldots, N)$. It is clear that P is maximized when all $\widehat{w}_i$ vanish except for $\widehat{w}_1$. If for such $\widehat{w}_i$ the condition $S \geq S_*$ is satisfied, then this is the optimal solution. However, if it is not satisfied, then we need to sacrifice P&L P in order to increase the Sharpe ratio S by allowing $\widehat{w}_2$, $\widehat{w}_3$, *etc.*, to be non-zero until (8) is satisfied, where $\widehat{\alpha}_i$ are sorted as follows.[10]

3.1.1 Sorting Alphas

For each subset

$$\widehat{w}_i > 0, \quad i = 1, \ldots, K \tag{16}$$

$$\widehat{w}_i = 0, \quad i = K+1, \ldots, N \tag{17}$$

the Sharpe ratio is given by (recall that $\widehat{\alpha}_i \geq 0$)

$$S(K) = \frac{\sum_{i=1}^{K} \widehat{\alpha}_i \ \widehat{w}_i}{\sqrt{\sum_{i=1}^{K} \widehat{w}_i^2}} \tag{18}$$

We want to find $\widehat{w}_i$, $i = 1, \ldots, K$ that maximize $S(K)$ subject to (14):

$$\sum_{i=1}^{K} \nu_i \ \widehat{w}_i = 1 \tag{19}$$

---

[10] This sorting is a relaxation algorithm: we start with the maximal P&L by allocating all investment into α_1 and at each successive step add one more alpha such that P&L is maximized when the weights are allocated to maximize the Sharpe ratio, until we reach the bound S_*.

Let us introduce a Lagrange multiplier μ:

$$\widetilde{S}(K) = S(K) + \mu \left(\sum_{i=1}^{K} \nu_i \, \widehat{w}_i - 1 \right) \tag{20}$$

Then the solution for $\widehat{w}_i$ that maximizes $S(K)$ subject to (19) is given by

$$\frac{\partial \widetilde{S}(K)}{\partial \widehat{w}_i} = 0 \tag{21}$$

$$\frac{\partial \widetilde{S}(K)}{\partial \mu} = 0 \tag{22}$$

The μ equation gives (19). The $\widehat{w}_i$ equations give

$$\widehat{w}_i = \gamma \, \widehat{\alpha}_i + \beta \, \nu_i, \quad i = 1, \dots, K \tag{23}$$

$$\gamma = \frac{\sum_{i=1}^{K} \widehat{w}_i^2}{\sum_{i=1}^{K} \widehat{\alpha}_i \, \widehat{w}_i} \tag{24}$$

$$\beta = \mu \, \gamma \, \left(\sum_{i=1}^{K} \widehat{w}_i^2 \right)^{\frac{1}{2}} \tag{25}$$

Plugging (23) into (24), we have two solutions, $\beta = -\gamma \sum_{i=1}^{K} \alpha_i \nu_i / \sum_{i=1}^{K} \nu_i^2$, and $\beta = 0$. However, the former would imply $\sum_{i=1}^{K} \nu_i \widehat{w}_i = 0$, so we have the following unique solution with $\beta = 0$ and $\mu = 0$:

$$\widehat{w}_i = \gamma \, \widehat{\alpha}_i \tag{26}$$

$$\gamma = \frac{1}{\sum_{i=1}^{K} \nu_i \, \widehat{\alpha}_i} \tag{27}$$

where γ is fixed using (19).

The corresponding P&L in Eq. (12) is given by

$$\frac{P(K)}{I} = \frac{\sum_{i=1}^{K} \widehat{\alpha}_i^2}{\sum_{i=1}^{K} \nu_i \, \widehat{\alpha}_i} = \frac{\sum_{i=1}^{K} \nu_i^2 \, \alpha_i^2}{\sum_{i=1}^{K} \nu_i^2 \, \alpha_i} \tag{28}$$

where we have used (9). In the remainder of this sub-subsection it will be more convenient to work with the original α_i as opposed to $\widehat{\alpha}_i$.

Let us sort α_i such that $P(K)$ is maximized for each K. I.e., for $K = 1$ we have α_1 such that $P(1)$ is maximized (which implies $\alpha_1 = \max(\alpha_i, i = 1 \dots, N)$), for $K = 2$ we have α_1 and α_2, where α_1 is fixed at $K = 1$ as above, and α_2 is such that $P(2)$ is maximized, *etc.* With this sorting, $P(K)$ is a monotonically decreasing function of K. To see this, let $\alpha_i, i = 1, \dots, K+2$ be sorted as above (*i.e.*, such

that at each K the P&L $P(K)$ is maximized), and let

$$a \equiv \sum_{i=1}^{K} \nu_i^2 \, \alpha_i \tag{29}$$

$$b \equiv \sum_{i=1}^{K} \nu_i^2 \, \alpha_i^2 \tag{30}$$

$$x \equiv \alpha_{K+1} \tag{31}$$

$$y \equiv \alpha_{K+2} \tag{32}$$

$$u \equiv \nu_{K+1}^2 \tag{33}$$

$$v \equiv \nu_{K+2}^2 \tag{34}$$

Without loss of generality, we will assume that all α_i are distinct, and all ν_i are distinct, $i = 1, \ldots, K + 2$. As above, let: $P(K)$ correspond to K alphas α_i, $i = 1, \ldots, K$; $P(K+1)$ correspond to $K+1$ alphas α_i, $i = 1, \ldots, K+1$; and $P(K+2)$ correspond to $K+2$ alphas α_i, $i = 1, \ldots, K+2$. Also, let $P'(K+1)$ correspond to $K+1$ alphas α_i', $i = 1, \ldots, K+1$, where $\alpha_i' = \alpha_i$ for $i = 1, \ldots, K$, and $\alpha_{K+1}' = \alpha_{K+2}$. By definition, we have

$$P(K+1) > P'(K+1) \tag{35}$$

This implies that

$$a \, u \, x^2 + v \, y \, \left(b + u \, x^2\right) > b \, u \, x + v \, y^2 \, (a + u \, x) \tag{36}$$

Now assume that $P(K+2) > P(K+1)$, which would imply

$$y > \frac{b + u \, x^2}{a + u \, x} \tag{37}$$

Together with (36) we would then have

$$x > \frac{b}{a} \tag{38}$$

and

$$P(K+1) - P(K) = \frac{b + u \, x^2}{a + u \, x} - \frac{b}{a} = \frac{u \, x \, (a \, x - b)}{a \, (a + u \, x)} > 0 \tag{39}$$

i.e., $P(K) < P(K+1) < P(K+2)$, and $P(K)$ must monotonically increase with K. However, this is not possible as $P(2) < P(1)$:

$$P(2) - P(1) = \nu_2^2 \, \alpha_2 \, \frac{\alpha_2 - \alpha_1}{\nu_1^2 \, \alpha_1 + \nu_2^2 \, \alpha_2} < 0 \tag{40}$$

as $\alpha_1 = \max(\alpha_i, i = 1 \ldots, N)$. This implies that for no K can $P(K+2) > P(K+1)$, and $P(K)$ monotonically decreases with K.

3.1.2 Computing Weights

With the aforementioned ordering of $\widehat{\alpha}_i$, for each subset

$$\widehat{w}_i > 0, \quad i = 1, \ldots, K \tag{41}$$

$$\widehat{w}_i = 0, \quad i = K+1, \ldots, N \tag{42}$$

the Sharpe ratio is maximized when (see above)

$$\widehat{w}_i = \gamma \, \widehat{\alpha}_i, \quad i = 1, \ldots, K \tag{43}$$

where $\gamma > 0$ is a constant. The corresponding Sharpe ratio is given by

$$S(K) = \sqrt{\sum_{i=1}^{K} \widehat{\alpha}_i^2} \tag{44}$$

The corresponding P&L in Eq. (12) is given by

$$\frac{P(K)}{I} = \frac{\sum_{i=1}^{K} \widehat{\alpha}_i^2}{\sum_{i=1}^{K} \nu_i \, \widehat{\alpha}_i} \tag{45}$$

Here, as above, we sort $\widehat{\alpha}_i$ such that $P(K)$ is maximized for each K. *I.e.*, for $K = 1$ we have $\widehat{\alpha}_1$ such that $P(1)$ is maximized, for $K = 2$ we have $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$, where $\widehat{\alpha}_1$ is fixed at $K = 1$ as above, and $\widehat{\alpha}_2$ is such that $P(2)$ is maximized, *etc.* As we showed above, with this sorting, $P(K)$ is a monotonically decreasing function of K.

Let K_* be such that

$$S(K_* - 1) < S_* \tag{46}$$

$$S(K_*) \geq S_* \tag{47}$$

Then a quasi-optimal solution (see below), for which we have $S = S_*$, is given by

$$\widehat{w}_i = \gamma \, \widehat{\alpha}_i, \quad i = 1, \ldots, K_* - 1 \tag{48}$$

$$\widehat{w}_{K_*} = \gamma \, \eta \, \widehat{\alpha}_{K_*} \tag{49}$$

$$\widehat{w}_i = 0, \quad i = K_* + 1, \ldots, N \tag{50}$$

where γ will be fixed via (14), and $0 < \eta \leq 1$ is fixed by the requirement that for these weights $S = S_*$:

$$\frac{\sum_{i=1}^{K_*-1} \widehat{\alpha}_i^2 + \eta \, \widehat{\alpha}_{K_*}^2}{\sqrt{\sum_{i=1}^{K_*-1} \widehat{\alpha}_i^2 + \eta^2 \, \widehat{\alpha}_{K_*}^2}} = S_* \tag{51}$$

which gives[11]

$$\eta = \frac{\widetilde{S}_1^2 - \widetilde{S}_1 \, \widetilde{S}_* \, \sqrt{\widetilde{S}_2^2 - \widetilde{S}_*^2}}{\widetilde{S}_*^2 - 1} \tag{52}$$

---

[11] Eq. (51) has two roots, and the correct root is fixed by the requirement that $0 < \eta \leq 1$.

where

$$\widetilde{S}_1 \equiv S(K_* - 1)/\widehat{\alpha}_{K_*} \tag{53}$$

$$\widetilde{S}_2 \equiv S(K_*)/\widehat{\alpha}_{K_*} \tag{54}$$

$$\widetilde{S}_* \equiv S_*/\widehat{\alpha}_{K_*} \tag{55}$$

$$S_2^2 = S_1^2 + 1 \tag{56}$$

Note that $\eta = 1$ if $S_* = S(K_*)$.

The reason why the above solution is not necessarily optimal is that $P(K)$ depends on the ordering of $\widehat{\alpha}_i$ – we order $\widehat{\alpha}_i$ such that at each step $K \to K + 1$ the P&L $P(K + 1)$ is maximized, but this does not guarantee that there is no other ordering for which the same Sharpe ratio S_* (or higher) cannot be achieved for a higher P&L, because the P&L can be path dependent. For large N this path dependence is expected to be suppressed.[12] In the next two subsections we discuss two ways of constructing the optimal solution.

3.2 Optimal Weights

In this subsection we discuss the problem of maximizing P&L with the Sharpe ratio fixed: $S = S_*$. We have

$$S = \frac{\sum_{i=1}^{N} \widehat{\alpha}_i \ \widehat{w}_i}{\sqrt{\sum_{i=1}^{N} \widehat{w}_i^2}} \tag{57}$$

The condition $S = S_*$ can be written as

$$\widehat{w}^T \ \Theta \ \widehat{w} = 0 \tag{58}$$

where

$$\Theta_{ij} \equiv \widehat{\alpha}_i \ \widehat{\alpha}_j - S_*^2 \ \delta_{ij} \tag{59}$$

The matrix Θ_{ij} has the following eigenvalue structure:

$$\Theta \ \phi = \theta_* \ \phi \tag{60}$$

$$\Theta \ \chi^{(A)} = \theta' \ \chi^{(A)}, \quad A = 1, \dots, N - 1 \tag{61}$$

where the eigenvectors ϕ and $\chi^{(A)}$ and the eigenvalues θ_* and θ' are given by

$$\phi_i = \frac{\widehat{\alpha}_i}{a} \tag{62}$$

$$\sum_{i=1}^{N} \chi_i^{(A)} \ \phi_i = 0 \tag{63}$$

$$\theta_* = a^2 - S_*^2 \tag{64}$$

$$\theta' = -S_*^2 \tag{65}$$

---

[12] More precisely, this is expected to be the case for generic alpha configurations.

and[13]

$$a \equiv \sqrt{\sum_{i=1}^{N} \widehat{\alpha}_i^2} \tag{66}$$

The eigenvectors $\chi^{(A)}$ are orthogonal to the eigenvector ϕ. In fact, let us take them to be orthonormal:

$$\sum_{i=1}^{N} \chi_i^{(A)} \chi_i^{(B)} = \delta_{AB} \tag{67}$$

where $A, B = 1, \ldots, N - 1$. Note that ϕ has norm 1 as defined above.

Note that the N eigenvectors ϕ and $\chi^{(A)}$ form a complete linearly independent set of N-vectors. Therefore, the optimal solution $\widehat{w}_i$ can be written as

$$\widehat{w}_i = \gamma \left(\phi_i + \sum_{A=1}^{N-1} \beta_A \chi_i^{(A)} \right) \tag{68}$$

where γ and β_A, $A = 1, \ldots, N - 1$ are N unknown coefficients we must determine. The P&L (12), the Sharpe ratio condition (58) and the weight normalization condition (14) now read:

$$\frac{P}{I} = a \, \gamma \tag{69}$$

$$\sum_{A=1}^{N-1} \beta_A^2 = r^2 \tag{70}$$

$$\gamma \left(\widetilde{\phi} + \sum_{A=1}^{N-1} \beta_A \, \widetilde{\chi}_A \right) = 1 \tag{71}$$

$$\phi_i + \sum_{A=1}^{N-1} \beta_A \, \chi_i^{(A)} \geq 0, \quad i = 1, \ldots, N \tag{72}$$

where the last condition follows from the requirement that for the optimal solution all $\widehat{w}_i \geq 0$ due to the fact that all $\widehat{\alpha}_i \geq 0$, and

$$r^2 \equiv -\frac{\theta_*}{\theta'} = \frac{a^2}{S_*^2} - 1 \tag{73}$$

$$\widetilde{\phi} \equiv \sum_{i=1}^{N} \nu_i \, \phi_i \tag{74}$$

$$\widetilde{\chi}_A \equiv \sum_{i=1}^{N} \nu_i \, \chi_i^{(A)} \tag{75}$$

---

[13] Note that $S_{\max} = a$, and $\widehat{w}_i = \widehat{\alpha}_i / \sum_{j=1}^{N} \nu_j \, \widehat{\alpha}_j$ for $S = S_{\max}$. Below we assume $S_* < S_{\max}$.

So, our optimization problem is now reduced to maximizing γ subject to (70), (71) and (72). *I.e.*, our optimization problem now reads:

$$Y \equiv \sum_{A=1}^{N-1} \beta_A \, \widetilde{\chi}_A \to \min \tag{76}$$

$$\sum_{A=1}^{N-1} \beta_A^2 = r^2 \tag{77}$$

$$\phi_i + \sum_{A=1}^{N-1} \beta_A \, \chi_i^{(A)} \geq 0, \quad i = 1, \ldots, N \tag{78}$$

This optimization problem can be solved as follows.

Note that β_A live on an $(N-2)$-sphere of radius r, so the sum in (76) is bounded both from below and above. Also, note that without loss of generality we can always assume that $\widetilde{\chi}_A$ has the following canonical form:

$$\widetilde{\chi}_1 \equiv \kappa > 0 \tag{79}$$
$$\widetilde{\chi}_A = 0, \quad A = 2, \ldots, N - 1 \tag{80}$$

We can always rotate $\chi_i^{(A)}$ and β_A into the above canonical form using an $SO(N-1)$ rotation (under which Y is invariant):

$$\chi_i^{(A)} \to \sum_{B=1}^{N-1} U_{AB} \, \chi_i^{(B)}, \quad A = 1, \ldots, N - 1 \tag{81}$$

$$\widetilde{\chi}_A \to \sum_{B=1}^{N-1} U_{AB} \, \widetilde{\chi}_B, \quad A = 1, \ldots, N - 1 \tag{82}$$

$$\beta_A \to \sum_{B=1}^{N-1} U_{AB} \, \beta_B, \quad A = 1, \ldots, N - 1 \tag{83}$$

$$\sum_{C=1}^{N-1} U_{AC} \, U_{BC} = \delta_{AB}, \quad A, B = 1, \ldots, N - 1 \tag{84}$$

and the sign of $\widetilde{\chi}_1$ can always be set by appropriately flipping the signs of the eigenvectors $\chi_i^{(A)}$, if need be.

In the above basis we have

$$Y = \kappa \, \beta_1 \tag{85}$$

Its maximum occurs at $\beta_1 = r$ (the "North Pole" on the $(N-2)$-sphere) and its minimum occurs at $\beta_1 = -r$ (the "South Pole" on the $(N-2)$-sphere). If $\beta_1 = -r$ satisfies the constraints (72), then this corresponds to the optimal solution. However, the constraints (72) may not be satisfied. From the quasi-optimal solution of Subsection 3.1 we know that a solution with $\widehat{w}_i \geq 0$ and the Sharpe ratio equal

S_* exists, *i.e.*, there exist values of β_A on the $(N-2)$-sphere (70) such that (72) are satisfied. However, this might not be the optimal solution. The optimal solution therefore lies along a path with decreasing β_1 connecting the point on the $(N-2)$-sphere corresponding to the quasi-optimal solution and the South Pole. (Note that all paths on an $(N-2)$-sphere are topologically equivalent.) Therefore, we need to identify this point corresponding to the optimal solution. One way of approaching this problem is via reducing the number of β_A to be determined by utilizing the residual $SO(N-2)$ rotational symmetry.

If for the optimal solution $-r < \beta_1 < r$, then at least one β_A, $A = 2, \ldots, N-1$ must be non-zero. Without loss of generality we can always assume that β_A have the following canonical form:

$$\beta_A = 0, \quad A = 3, \ldots, N-1 \tag{86}$$

We can always rotate $\chi_i^{(A)}$ and β_A, $A = 2, \ldots, N-1$ into the above canonical form[14] using an $SO(N-2)$ rotation (under which $\widetilde{\chi}_A = 0$, $A = 2, \ldots, N-1$ are invariant):

$$\chi_i^{(A)} \to \sum_{B=2}^{N-1} U'_{AB} \, \chi_i^{(B)}, \quad A = 2, \ldots, N-1 \tag{87}$$

$$\beta_A \to \sum_{B=2}^{N-1} U'_{AB} \, \beta_B, \quad A = 2, \ldots, N-1 \tag{88}$$

$$\sum_{C=2}^{N-1} U'_{AC} \, U'_{BC} = \delta_{AB}, \quad A, B = 2, \ldots, N-1 \tag{89}$$

So we have:

$$\phi_i + \beta_1 \, \chi_i^{(1)} + \beta_2 \, \chi_i^{(2)} \geq 0, \quad i = 1, \ldots, N \tag{90}$$
$$\beta_1^2 + \beta_2^2 = r^2 \tag{91}$$

Also, since ϕ_i and $\chi_i^{(A)}$, $A = 1, \ldots, N-1$ form a complete orthonormal set of N-vectors, there exist coefficients c and d_A, $A = 1, \ldots, N-1$ such that

$$c \, \phi_i + \sum_{A=1}^{N-1} d_A \, \chi_i^{(A)} = \nu_i, \quad i = 1, \ldots, N \tag{92}$$

---

[14] Note, however, that while we are free to choose any basis for $\chi_i^{(A)}$, assuming the above canonical form implies that $\chi_i^{(2)}$ is now unknown and must be determined rather than chosen. *I.e.*, instead of choosing $\chi_i^{(A)}$ and trying to determine β_A, we restrict β_A to the above canonical form at the expense of having to determine $\chi_i^{(2)}$.

which implies that

$$c = \widetilde{\phi} \tag{93}$$

$$d_1 = \kappa \tag{94}$$

$$d_A = 0, \quad A = 2, \dots, N-1 \tag{95}$$

$$\chi_i^{(1)} = \frac{1}{\kappa} \left(\nu_i - \widetilde{\phi} \, \phi_i \right) \tag{96}$$

$$\kappa^2 = \widetilde{\nu}^2 - \widetilde{\phi}^2 \tag{97}$$

where

$$\widetilde{\nu}^2 \equiv \sum_{i=1}^{N} \nu_i^2 \tag{98}$$

We have the following constraints:

$$\chi_i^{(2)} \geq -\frac{1}{\kappa \sqrt{r^2 - \beta_1^2}} \left(\beta_1 \, \nu_i + \left[\kappa - \beta_1 \, \widetilde{\phi} \right] \phi_i \right) \tag{99}$$

where without loss of generality we assume that $\beta_2 > 0$. (We discuss the $\beta_2 < 0$ case below.)

There are only four *a priori* conditions (which are the $\widetilde{\chi}_2 = 0$ condition plus the relevant orthonormality conditions) that the constraints (99) must be compatible with:

$$\sum_{i=1}^{N} \nu_i \, \chi_i^{(2)} = 0 \tag{100}$$

$$\sum_{i=1}^{N} \phi_i \, \chi_i^{(2)} = 0 \tag{101}$$

$$\sum_{i=1}^{N} \chi_i^{(1)} \, \chi_i^{(2)} = 0 \tag{102}$$

$$\sum_{i=1}^{N} \left[\chi_i^{(2)} \right]^2 = 1 \tag{103}$$

Due to (96), (102) follows from (100) and (101). Also, (101) is automatically compatible with (99). A nontrivial condition follows from (100) and (99) via multiplying both sides of (99) by ν_i and summing over $i = 1, \dots, N$:

$$\widetilde{\nu}^2 \, \beta_1 + \widetilde{\phi} \left[\kappa - \beta_1 \, \widetilde{\phi} \right] \geq 0 \tag{104}$$

which gives

$$\beta_1 \geq \beta_* \equiv -\frac{\widetilde{\phi}}{\kappa} \tag{105}$$

12

Considering that $\beta_1 \geq -r$, this condition is satisfied if

$$\widetilde{\phi} \geq \widetilde{\nu} \sqrt{1 - \frac{S_*^2}{a^2}} \tag{106}$$

If this condition is not satisfied, then $\beta_1 > -r$ and the optimal solution satisfying (72) lies away from the South Pole. Note that (106) is a necessary condition to have $\beta_1 = -r$.

Thus, our optimization problem has been reduced to finding the lowest value of β_1 such that there exits an N-vector $\chi^{(2)}$ such that (99), (100), (101) and (103) are satisfied. Assuming $\beta_1 \geq \beta_*$, the constraints (99) can be incompatible with (103), which further constrains β_1. For a given β_1 let T^+ and T^- be the subsets of the values of the index $i = 1, \ldots, N$ such that

$$\phi_i < -\frac{\beta_1}{\kappa - \beta_1 \, \widetilde{\phi}} \, \nu_i \, , \quad i \in T^+ \tag{107}$$

$$\phi_i \geq -\frac{\beta_1}{\kappa - \beta_1 \, \widetilde{\phi}} \, \nu_i \, , \quad i \in T^- \tag{108}$$

Then we have $\chi_i^{(2)} > 0$ for $i \in T^+$ and (103) may or may not be attainable for such β_1. In the following we will treat T^+ and T^- as vectors of lengths N^+ and N^- ($N^+ + N^- = N$), and for $i \in T^-$ we will use a map $i = T_a^-$, where $a = 1, \ldots, N^-$.

We can construct $\chi^{(2)}$ as follows:

$$\chi_i^{(2)} = -\frac{\theta_i}{\kappa \, \sqrt{r^2 - \beta_1^2}} \, \left(\beta_1 \, \nu_i + \left[\kappa - \beta_1 \, \widetilde{\phi} \right] \phi_i \right), \quad i \in T^+ \tag{109}$$

$$\theta_i \geq 1, \quad i \in T^+ \tag{110}$$

$$\sum_{i \in T^+} \left[\chi_i^{(2)} \right]^2 \equiv \widetilde{\zeta} \tag{111}$$

$$\sum_{i \in T^+} \nu_i \, \chi_i^{(2)} \equiv -\widetilde{\chi} \tag{112}$$

$$\sum_{i \in T^+} \chi_i^{(2)} \, \phi_i \equiv -\widetilde{\xi} \tag{113}$$

$$\chi_a^- \equiv \chi_i^{(2)}, \quad i = T_a^-, \quad a = 1, \ldots, N^- \tag{114}$$

$$\varphi_a \equiv \phi_i, \quad i = T_a^-, \quad a = 1, \ldots, N^- \tag{115}$$

$$\omega_a \equiv \nu_i, \quad i = T_a^-, \quad a = 1, \ldots, N^- \tag{116}$$

The N^--vector χ_a^- is constrained as follows:

$$\sum_{a=1}^{N^-} \omega_a \, \chi_a^- = \widetilde{\chi} \tag{117}$$

$$\sum_{a=1}^{N^-} \chi_a^- \, \varphi_a = \widetilde{\xi} \tag{118}$$

$$\sum_{a=1}^{N^-} \left[\chi_a^-\right]^2 = 1 - \widetilde{\zeta} \tag{119}$$

The vector χ_a^- can always be decomposed as follows:

$$\chi_a^- = c_- \, \varphi_a + d_- \, \omega_a + \widehat{\chi}_a^- \tag{120}$$

$$c_- \equiv \frac{\widetilde{\lambda} \, \widetilde{\xi} - \widetilde{\varphi} \, \widetilde{\chi}}{\widetilde{\sigma} \, \widetilde{\lambda} - \widetilde{\varphi}^2} \tag{121}$$

$$d_- \equiv \frac{\widetilde{\sigma} \, \widetilde{\chi} - \widetilde{\varphi} \, \widetilde{\xi}}{\widetilde{\sigma} \, \widetilde{\lambda} - \widetilde{\varphi}^2} \tag{122}$$

where

$$\widetilde{\sigma} \equiv \sum_{a=1}^{N^-} \varphi_a^2 \tag{123}$$

$$\widetilde{\varphi} \equiv \sum_{a=1}^{N^-} \omega_a \, \varphi_a \tag{124}$$

$$\widetilde{\lambda} \equiv \sum_{a=1}^{N^-} \omega_a^2 \tag{125}$$

and the N^--vector $\widehat{\chi}_a^-$ is such that

$$\sum_{a=1}^{N^-} \widehat{\chi}_a^- \, \omega_a = 0 \tag{126}$$

$$\sum_{a=1}^{N^-} \widehat{\chi}_a^- \, \varphi_a = 0 \tag{127}$$

With this decomposition, (119) reads

$$\sum_{a=1}^{N^-} \left[\widehat{\chi}_a^-\right]^2 = 1 - \widetilde{\zeta} - \frac{g(\widetilde{\chi}, \widetilde{\xi})}{\widetilde{\sigma} \, \widetilde{\lambda} - \widetilde{\varphi}^2} \tag{128}$$

$$g(x_1, x_2) \equiv \widetilde{\sigma} \, x_1^2 + \widetilde{\lambda} \, x_2^2 - 2 \, \widetilde{\varphi} \, x_1 \, x_2 \tag{129}$$

14

Note that

$$\widetilde{\sigma} = \sum_{i \in T^-} \phi_i^2 = \frac{\sum_{i \in T^-} \widehat{\alpha}_i^2}{\sum_{i=1}^N \widehat{\alpha}_i^2} \leq 1 \tag{130}$$

$$\frac{\widetilde{\varphi}^2}{\widetilde{\sigma}} = \frac{\left(\sum_{i \in T^-} \nu_i \, \widehat{\alpha}_i\right)^2}{\sum_{i \in T^-} \widehat{\alpha}_i^2} < \widetilde{\lambda} = \sum_{i \in T^-} \nu_i^2 \tag{131}$$

where we assume that $\widehat{\alpha}_i$, $i \in T^-$ are not all identical, and ν_i, $i \in T^-$ are not all identical. Further, note that each of $\widetilde{\zeta}$, $\widetilde{\xi}$ and $\widetilde{\chi}$ is minimized when $\theta_i \equiv 1$, $i \in T^+$, while if we fix θ_i, then $\widetilde{\zeta}$ increases as β_1 decreases (assuming $\beta_1 < 0$). Finally, the eigenvalues of the 2×2 curvature matrix

$$\frac{\partial^2 g(x_1, x_2)}{\partial x_p \, \partial x_q}, \quad p, q = 1, 2 \tag{132}$$

are given by

$$g_\pm = \frac{1}{2} \left(\widetilde{\lambda} + \widetilde{\sigma} \pm \sqrt{\left(\widetilde{\lambda} - \widetilde{\sigma}\right)^2 + 4 \, \widetilde{\varphi}^2} \right) \tag{133}$$

and are both positive considering that $\widetilde{\varphi}^2 < \widetilde{\sigma} \, \widetilde{\lambda}$. This implies that the r.h.s. of (128) is maximized when $\theta_i \equiv 1$, $i \in T^+$. Therefore, the above construction provides $\chi_i^{(2)}$ corresponding to the optimal (*i.e.*, minimal) β_1 when $\theta_i \equiv 1$, $i \in T^+$ and $\widehat{\chi}_a^- \equiv 0$, $a = 1, \ldots, N^-$:

$$\chi_i^{(2)} = -\frac{1}{\kappa \, \sqrt{r^2 - \beta_1^2}} \left(\beta_1 \, \nu_i + \left[\kappa - \beta_1 \, \widetilde{\phi} \right] \phi_i \right), \quad i \in T^+ \tag{134}$$

$$\chi_i^{(2)} = c_- \, \phi_i + d_- \, \nu_i, \quad i \in T^- \tag{135}$$

$$\widetilde{\zeta} + \frac{g(\widetilde{\chi}, \widetilde{\xi})}{\widetilde{\sigma} \, \widetilde{\lambda} - \widetilde{\varphi}^2} = 1 \tag{136}$$

The last equation (136) is then used to determine the optimal value of β_1. Practically, one needs to employ an iterative procedure to determine β_1 from (136) because T^+ and T^- depend on β_1. One way of implementing such an iterative procedure is to use the quasi-optimal solution of the previous subsection. Let $z_i \equiv \widehat{w}_i$ for that solution. Then the corresponding value of $\beta_1 \equiv \widehat{\beta}_1$ is given by (this follows from (69), (71), (79) and (80))

$$\widehat{\beta}_1 = \frac{1 - \widetilde{\phi} \, \sum_{i=1}^N z_i \, \phi_i}{\kappa \, \sum_{i=1}^N z_i \, \phi_i} \tag{137}$$

We can now search for the value of β_1 between $\widehat{\beta}_1$ and β_*, *e.g.*, using a standard algorithm such as successively testing midpoints until the l.h.s. of (136) approaches 1 from below with a desired precision.

Above we assume that $\beta_2 > 0$. If we assume that $\beta_2 < 0$, then instead of (99) we have:

$$\chi_i^{(2)} \leq \frac{1}{\kappa \sqrt{r^2 - \beta_1^2}} \left(\beta_1 \, \nu_i + \left[\kappa - \beta_1 \, \widetilde{\phi} \right] \phi_i \right) \tag{138}$$

and the above discussion carries through unchanged except that now $\chi_i^{(2)}$, $i \in T^+$ are all negative.

Finally, the weights $\widehat{w}_i = 0$ for $i \in T^+$, while for $i \in T^-$ they are given by

$$\widehat{w}_i = \frac{\left(1 - \frac{\beta_1}{\kappa} \, \widetilde{\phi} + \sqrt{r^2 - \beta_1^2} \, c_- \right) \phi_i + \left(\frac{\beta_1}{\kappa} + \sqrt{r^2 - \beta_1^2} \, d_- \right) \nu_i}{\widetilde{\phi} + \beta_1 \, \kappa} \tag{139}$$

The fact that non-zero $\widehat{w}_i$ have a form $\widehat{w}_i = e \, \phi_i + f \, \nu_i$ (where e and f are coefficients) is not surprising from symmetry considerations: ϕ_i and ν_i are the only vectors available as building blocks. However, what is nontrivial is the definition of T^+ and T^-, *i.e.*, which $\widehat{w}_i$ are vanishing.

3.3 Alternative Construction

We can construct the above optimal solution in a different way. Instead of maximizing P&L directly, we can i) first minimize the volatility R (*i.e.*, maximize the Sharpe ratio S) for a fixed value of the P&L P, and ii) then find P such that $S = S_*$. *I.e.*, at the first step we solve the following problem:

$$\frac{P}{I} = \sum_{i=1}^{N} \widehat{\alpha}_i \, \widehat{w}_i \equiv \widetilde{P} = \text{fixed} \tag{140}$$

$$\frac{R}{I} = \sqrt{\sum_{i=1}^{N} \widehat{w}_i^2} \to \min \tag{141}$$

$$\sum_{i=1}^{N} \nu_i \, \widehat{w}_i = 1 \tag{142}$$

$$\widehat{w}_i \geq 0, \quad i = 1, \dots, N \tag{143}$$

This problem can be stated as follows:

$$g(\widehat{w}, \mu, \widetilde{\mu}) \equiv \frac{1}{2} \sum_{i=1}^{N} \widehat{w}_i^2 + \mu \left(\sum_{i=1}^{N} \nu_i \, \widehat{w}_i - 1 \right) + \widetilde{\mu} \left(\sum_{i=1}^{N} \widehat{\alpha}_i \, \widehat{w}_i - \widetilde{P} \right) \tag{144}$$

$$g(\widehat{w}, \mu, \widetilde{\mu}) \to \min \tag{145}$$

$$\widehat{w}_i \geq 0, \quad i = 1, \dots, N \tag{146}$$

where μ and $\widetilde{\mu}$ are Lagrange multipliers, *i.e.*, the objective function $g(\widehat{w}, \mu, \widetilde{\mu})$ is minimized w.r.t. $\widehat{w}_i$, μ and $\widetilde{\mu}$.

16

The solution is as follows. First, let us sort α_i (*not* $\widehat{\alpha}_i$) in the decreasing order (recall that $\widehat{\alpha}_i = \nu_i \, \alpha_i$), *i.e.*, $\alpha_1 = \max(\alpha_i, i = 1, \ldots, N)$. Then

$$\widehat{w}_i = \frac{\left(\widetilde{\nu}^2 \, \widetilde{P} - a\right)\widehat{\alpha}_i - \left(a \, \widetilde{P} - b^2\right)\nu_i}{\widetilde{\nu}^2 \, b^2 - a^2}, \quad i = 1, \ldots, K \tag{147}$$

$$\widehat{w}_i = 0, \quad i = K + 1, \ldots, N \tag{148}$$

$$a \equiv \sum_{i=1}^{K} \nu_i \, \widehat{\alpha}_i \tag{149}$$

$$b^2 \equiv \sum_{i=1}^{K} \widehat{\alpha}_i^2 \tag{150}$$

$$\widetilde{\nu}^2 \equiv \sum_{i=1}^{K} \nu_i^2 \tag{151}$$

$$\widehat{\alpha}_i > \frac{a \, \widetilde{P} - b^2}{\widetilde{\nu}^2 \, \widetilde{P} - a} \, \nu_i, \quad i = 1, \ldots, K \tag{152}$$

where the last condition determines K, *i.e.*, K is the maximum number of first K alphas α_i (sorted in the decreasing order) such that (152), which is equivalent to

$$\alpha_i > \frac{a \, \widetilde{P} - b^2}{\widetilde{\nu}^2 \, \widetilde{P} - a}, \quad i = 1, \ldots, K \tag{153}$$

is satisfied. Note that $\widetilde{\nu}^2 \, b^2 > a^2$ (assuming non-identical $\widehat{\alpha}_i$ and non-identical ν_i). Also, $\widetilde{P} \leq \alpha_1$, and we can assume $\widetilde{P} > 0$. Finally, note that $\widetilde{P} = b^2/a$ corresponds to

$$S = S_{\mathrm{max}} \equiv \sqrt{\sum_{i=1}^{N} \widehat{\alpha}_i^2} \tag{154}$$

Indeed, in this case we have $\widehat{w}_i = \widehat{\alpha}_i/a$ (and all $\widehat{w}_i$ are non-vanishing unless some $\widehat{\alpha}_i = 0$). Therefore, we can assume that $\widetilde{P} \geq b^2/a$, because, if we consider lower $\widetilde{P}$, we have a solution with higher Sharpe ratio and higher P&L, so the search can be limited to the values $b^2/a < \widetilde{P} < \alpha_1$.

Using the above algorithm we can construct $\widehat{w}_i$ for any given $\widetilde{P}$. We can then determine the weights corresponding to $S = S_*$ using a standard algorithm such as successively testing midpoints between successive values of $\widetilde{P}$ (between α_1 and b^2/a) until the Sharpe ratio S approaches S_* with a desired precision.

4   Non-Diagonal Case

When the covariance matrix C_{ij} is non-diagonal, things are more complicated because of the modulus in (3). We can still go into a diagonal basis via the following

trick. Let $V_i^{(a)}$ be N right eigenvectors of C_{ij} corresponding to its eigenvalues $\lambda^{(a)}$, $a = 1, \ldots, N$:

$$C \ V^{(a)} = \lambda^{(a)} \ V^{(a)} \tag{155}$$

with no summation over a. Since C is symmetric, $V^{(a)}$ can be chosen to be orthonormal:

$$\sum_{i=1}^{N} V_i^{(a)} \ V_i^{(b)} = \delta_{ab} \tag{156}$$

which we assume to be the case. We will also assume that all $\lambda^{(a)} > 0$, *i.e.*, C_{ij} is positive-definite.[15] Furthermore, we will assume that none of the eigenvalues $\lambda^{(a)}$ is "infinitesimally" small compared with others.[16]

Now consider N synthetic portfolios $\Pi^{(a)}$, $a = 1, \ldots, N$, with the weights

$$w_i^{(a)} = \frac{V_i^{(a)}}{\sum_{j=1}^{N} \left| V_j^{(a)} \right|} \tag{157}$$

Note that

$$\sum_{i=1}^{N} \left| w_i^{(a)} \right| = 1 \tag{158}$$

Since we are assuming that all alphas are traded on the same execution platform, the portfolios $\Pi^{(a)}$ are tradable. Furthermore, the covariance matrix for these portfolios is diagonal:

$$C_{ab} \equiv \left\langle \alpha^{(a)}, \alpha^{(b)} \right\rangle = \frac{\lambda^{(a)}}{\left(\sum_{j=1}^{N} \left| V_j^{(a)} \right| \right)^2} \ \delta_{ab} \tag{159}$$

The corresponding alphas are given by

$$\alpha^{(a)} \equiv \frac{1}{\sum_{j=1}^{N} \left| V_j^{(a)} \right|} \sum_{i=1}^{N} V_i^{(a)} \ \alpha_i \tag{160}$$

and can be assumed to be non-negative – if any $\alpha^{(a)}$ is negative, we can always make it positive by flipping the signs of the weights in the corresponding synthetic portfolio. Now we could proceed to solve the optimization problem for these synthetic alphas as we did in the previous section. However, the caveat is that if we invest I dollars into a combinations of these synthetic $\alpha^{(a)}$, it does not correspond to investing I dollars into the underlying alphas due to "netting" of alphas[17] as the

---

[15] A simple method for making a covariance matrix positive-definite was discussed in [58] based on [59].

[16] *I.e.*, none of these eigenvalues are zeros distorted by machine precision, as would be the case if $M < N$ (where $M + 1$ is the number of observations in the alpha time series – see Section 2).

[17] Not to be confused with the "netting" of underlying tradables.

weights $w_i^{(a)}$ in various portfolios have opposite signs. So, in the optimization problem, on paper, we need to invest some *priori* unknown "synthetic" amount I' into the synthetic alphas $\alpha^{(a)}$ to have the desired actual investment I in the real alphas α_i. This amount I' would then have to be determined via an iterative procedure. However, for general C_{ij} this iterative procedure can run into stability issues. This is because the optimization now becomes effectively 2-dimensional. In this regard, it is more streamlined to tackle the problem by dealing with the modulus in (3) directly as opposed to attempting to circumvent it via diagonalization. We discuss this approach in the next section.

5 Optimal Weights

Here we will follow the framework of Subsection 3.3 and i) first minimize the volatility R (*i.e.*, maximize the Sharpe ratio S) for a fixed value of the P&L P, and ii) then find P such that $S = S_*$. The second step is straightforward.[18] At the first step we solve the following problem:

$$\frac{P}{I} = \sum_{i=1}^{N} \alpha_i \ w_i \equiv \widetilde{P} = \text{fixed} \tag{161}$$

$$\frac{R}{I} = \sqrt{\sum_{i,j=1}^{N} C_{ij} \ w_i \ w_j} \equiv \widetilde{R} \to \min \tag{162}$$

subject to

$$\sum_{i=1}^{N} |w_i| = 1 \tag{163}$$

where C_{ij} is arbitrary except that, without loss of generality for our purposes here (see below), we will assume it to be positive-definite. This problem can be stated as follows:

$$g(w, \mu, \widetilde{\mu}) \equiv \frac{1}{2} \sum_{i,j=1}^{N} C_{ij} \ w_i \ w_j +$$

$$+ \mu \left(\sum_{i=1}^{N} |w_i| - 1 \right) + \widetilde{\mu} \left(\sum_{i=1}^{N} \alpha_i \ w_i - \widetilde{P} \right) \to \min \tag{164}$$

where μ and $\widetilde{\mu}$ are Lagrange multipliers, *i.e.*, the objective function $g(w, \mu, \widetilde{\mu})$ is minimized w.r.t. w_i, μ and $\widetilde{\mu}$. What complicates matters is the modulus of w_i. The problem can still be solved, albeit it requires a finite iterative procedure, *i.e.*, the solution is exact and is obtained after a finite number of iterations.[19]

---

[18] The second step is a one-dimensional root searching problem.
[19] More precisely, the optimum is unique for $\mu \geq 0$ – see below.

Let J and J' be the subsets of the index $i = 1, \ldots, N$ such that

$$w_i \neq 0, \quad i \in J \tag{165}$$

$$w_i = 0, \quad i \in J' \tag{166}$$

Let

$$\eta_i \equiv \operatorname{sign}(w_i), \quad i \in J \tag{167}$$

Note that, since the modulus has a discontinuous derivative, the minimization equations are not the same as setting first derivatives of $g(w, \mu, \widetilde{\mu})$ to zero. More concretely, first derivatives are well-defined for $i \in J$, but not for $i \in J'$. So, we have the following minimization equations for w_i, $i \in J$:

$$\sum_{j \in J} C_{ij} \, w_j + \mu \, \eta_i + \widetilde{\mu} \, \alpha_i = 0, \quad i \in J \tag{168}$$

$$\sum_{i \in J} |w_i| = 1 \tag{169}$$

$$\sum_{i \in J} \alpha_i \, w_i = \widetilde{P} \tag{170}$$

with μ and $\widetilde{\mu}$ determined using (169) and (170). There are additional conditions for the global minimum corresponding to the directions $i \in J'$:

$$\sum_{i,j=1}^{N} C_{ij} \, (w_i + \epsilon_i) \, (w_j + \epsilon_j) \geq \sum_{i,j \in J} C_{ij} \, w_i \, w_j \tag{171}$$

$$\sum_{i=1}^{N} |w_i + \epsilon_i| = 1 \tag{172}$$

$$\sum_{i=1}^{N} \alpha_i \, (w_i + \epsilon_i) = \widetilde{P} \tag{173}$$

where w_i, $i \in J$ are determined using (168), while $w_i = 0$, $i \in J'$. The conditions (171) must be satisfied including for arbitrary infinitesimal ϵ_i subject to (172) and (173). For infinitesimal ϵ_i these conditions can be rewritten as follows:[20]

$$\sum_{i,j \in J} C_{ij} \, w_i \, \epsilon_j + \sum_{i \in J} \sum_{j \in J'} C_{ij} \, w_i \, \epsilon_j \geq 0 \tag{174}$$

$$\sum_{i \in J} \eta_i \, \epsilon_i + \sum_{i \in J'} |\epsilon_i| = 0 \tag{175}$$

$$\sum_{i \in J} \alpha_i \, \epsilon_i + \sum_{i \in J'} \alpha_i \, \epsilon_i = 0 \tag{176}$$

---

[20] Since here ϵ_i are taken to be infinitesimal, these are the conditions for a local minimum. See Subsection 5.5 for the global minimum conditions.

which, taking into account (168), reduce to

$$\sum_{j \in J'} \left(\sum_{i \in J} C_{ij} \ w_i \ \epsilon_j + \mu \ |\epsilon_j| + \widetilde{\mu} \ \alpha_j \ \epsilon_j \right) \geq 0 \qquad (177)$$

Since ϵ_j, $j \in J'$ are arbitrary, this gives the following conditions:

$$\forall j \in J' : \quad \left| \widetilde{\mu} \ \alpha_j + \sum_{i \in J} C_{ij} \ w_i \right| \leq \mu \qquad (178)$$

These conditions must be satisfied by the solution to (168), (169) and (170). The solution that (locally) minimizes $g(w, \mu, \widetilde{\mu})$ is given by (in matrix notation)

$$w = -\mu \ D \ \eta - \widetilde{\mu} \ D \ \alpha \qquad (179)$$

$$\mu = \frac{\widetilde{P} \ (\alpha^T \ D \ \eta) - (\alpha^T \ D \ \alpha)}{(\alpha^T \ D \ \alpha) \ (\eta^T \ D \ \eta) - (\alpha^T \ D \ \eta)^2} \qquad (180)$$

$$\widetilde{\mu} = \frac{(\alpha^T \ D \ \eta) - \widetilde{P} \ (\eta^T \ D \ \eta)}{(\alpha^T \ D \ \alpha) \ (\eta^T \ D \ \eta) - (\alpha^T \ D \ \eta)^2} \qquad (181)$$

Here all (both contracted and free) indices are assumed to run over $i \in J$, e.g.,

$$w_i = -\mu \sum_{j \in J} D_{ij} \ \eta_j - \widetilde{\mu} \sum_{j \in J} D_{ij} \ \alpha_j, \quad i \in J \qquad (182)$$

and D is the inverse matrix of the $N(J) \times N(J)$ matrix C_{ij}, $i, j \in J$, where $N(J) \equiv |J|$ is the number of elements of J:

$$\sum_{k \in J} C_{ik} \ D_{kj} = \delta_{ij}, \quad i, j \in J \qquad (183)$$

i.e., D is *not* a restriction of the inverse of the $N \times N$ matrix C_{ij}, $i, j \in 1, \dots, N$ to $i, j \in J$.

Here the following observation is in order. In the above solution, *a priori* we do not know i) what the subset J' is and ii) what the values of η_i are for $i \in J$. This means that *a priori* we have total of 3^N possible combinations (including the redundant empty J case), so if we go through this finite set, we will solve the problem exactly. However, 3^N is a prohibitively large number,[21] so one needs a more clever way of solving the problem.[22]

---

[21] We assume that N itself is large.

[22] We need to iterate because of the term proportional to μ in (182). Let us assume $\mu = 0$. Then for generic alpha configurations J' is empty, $D = C^{-1}$, and the solution is $w = (\alpha^T C^{-1} \eta)^{-1} C^{-1} \alpha$ with $\widetilde{P} = (\alpha^T C^{-1} \alpha)/(\alpha^T C^{-1} \eta)$, which corresponds to the solution to $S \to$ max with $S = S_{\max} = \sqrt{\alpha^T C^{-1} \alpha}$. We discuss this case in more detail in Subsection 5.3 in the context of the factor model covariance matrix.

5.1 Factor Model

As was pointed out in [57] in the context of optimization with linear costs, a way to reduce the number of iterations in this types of problems is to reduce the "off-diagonality" of C_{ij}. This can be achieved by considering a factor model for alphas.[23]

As in the case of a stock multi-factor risk model, instead of N alphas, one deals with $F \ll N$ risk factors and the covariance matrix C_{ij} is replaced by Γ_{ij} given by

$$\Gamma \equiv \Xi + \Omega \; \Phi \; \Omega^T \equiv \Xi + \widetilde{\Omega} \; \widetilde{\Omega}^T \tag{184}$$

$$\Xi_{ij} \equiv \xi_i^2 \; \delta_{ij} \tag{185}$$

$$\widetilde{\Phi} \; \widetilde{\Phi}^T = \Phi \tag{186}$$

where ξ_i is the specific risk for each α_i; Ω_{iA} is an $N \times F$ factor loadings matrix; $\widetilde{\Omega} \equiv \Omega \; \widetilde{\Phi}$; Φ_{AB} is the factor covariance matrix, $A, B = 1, \ldots, F$; and $\widetilde{\Phi}_{AB}$ is the Cholesky decomposition of Φ_{AB}, which is assumed to be positive-definite. *I.e.*, the random processes Υ_i corresponding to N alphas are modeled via N random processes z_i (corresponding to specific risk) together with F random processes f_A (corresponding to factor risk):

$$\Upsilon_i = z_i + \sum_{A=1}^{F} \Omega_{iA} \; f_A \tag{187}$$

$$\langle z_i, z_j \rangle = \Xi_{ij} \tag{188}$$

$$\langle z_i, f_A \rangle = 0 \tag{189}$$

$$\langle f_A, f_B \rangle = \Phi_{AB} \tag{190}$$

$$\langle \Upsilon_i, \Upsilon_j \rangle = \Gamma_{ij} \tag{191}$$

Instead of an $N \times N$ covariance matrix C_{ij} we now have an $F \times F$ covariance matrix Φ_{AB}. In the following we assume that C_{ij} has the factor model form Γ_{ij}.

5.2 Optimization with Factor Model

In this framework, the problem reduces to solving an $(F+2)$-dimensional system as follows. First, let

$$v_A \equiv \sum_{i=1}^{N} w_i \; \widetilde{\Omega}_{iA} = \sum_{i \in J} w_i \; \widetilde{\Omega}_{iA}, \quad A = 1, \ldots, F \tag{192}$$

Then from (168) we have

$$w_i = -\frac{1}{\xi_i^2} \left(\mu \; \eta_i + \widetilde{\mu} \; \alpha_i + \sum_{A=1}^{F} \widetilde{\Omega}_{iA} \; v_A \right), \quad i \in J \tag{193}$$

---

[23] Here we simply assume a factor model form for the covariance matrix without delving into details of how it is constructed.

Recalling that $w_i \, \eta_i > 0$, $i \in J$, and *assuming* that $\mu \geq 0$ (see below), we get

$$\eta_i = -\text{sign}\left(\widetilde{\mu} \, \alpha_i + \sum_{A=1}^{F} \widetilde{\Omega}_{iA} \, v_A\right), \quad i \in J \tag{194}$$

$$\forall i \in J: \quad \left|\widetilde{\mu} \, \alpha_i + \sum_{A=1}^{F} \widetilde{\Omega}_{iA} \, v_A\right| > \mu \tag{195}$$

$$\forall i \in J': \quad \left|\widetilde{\mu} \, \alpha_i + \sum_{A=1}^{F} \widetilde{\Omega}_{iA} \, v_A\right| \leq \mu \tag{196}$$

where (195) follows from (193), and (196) follows from (178). The last two inequalities define J and J' in terms of $(F+2)$ unknowns v_A, μ and $\widetilde{\mu}$.

In the remainder of this subsection we assume that $\mu \geq 0$. We discuss the $\mu < 0$ case in Subsection 5.4.

Substituting (193) into (192), (169) and (170), we get the following system of $(F+2)$ equations for $(F+2)$ unknowns v_A, μ and $\widetilde{\mu}$:

$$\sum_{B=1}^{F} Q_{AB} \, v_B + a_A \, \mu + b_A \, \widetilde{\mu} = 0 \tag{197}$$

$$\sum_{A=1}^{F} a_A \, v_A + c \, \mu + d \, \widetilde{\mu} = -1 \tag{198}$$

$$\sum_{A=1}^{F} b_A \, v_A + d \, \mu + e \, \widetilde{\mu} = -\widetilde{P} \tag{199}$$

where

$$Q_{AB} \equiv \delta_{AB} + \sum_{i \in J} \frac{\widetilde{\Omega}_{iA} \, \widetilde{\Omega}_{iB}}{\xi_i^2} \tag{200}$$

$$a_A \equiv \sum_{i \in J} \frac{\eta_i \, \widetilde{\Omega}_{iA}}{\xi_i^2} \tag{201}$$

$$b_A \equiv \sum_{i \in J} \frac{\alpha_i \, \widetilde{\Omega}_{iA}}{\xi_i^2} \tag{202}$$

$$c \equiv \sum_{i \in J} \frac{1}{\xi_i^2} \tag{203}$$

$$d \equiv \sum_{i \in J} \frac{\alpha_i \, \eta_i}{\xi_i^2} \tag{204}$$

$$e \equiv \sum_{i \in J} \frac{\alpha_i^2}{\xi_i^2} \tag{205}$$

23

Let $\widetilde{Q}_{ab}$, $a, b = 1, \ldots, (F+2)$ be the following symmetric $(F+2) \times (F+2)$ matrix:

$$\widetilde{Q}_{AB} \equiv Q_{AB} \tag{206}$$

$$\widetilde{Q}_{A,(F+1)} = \widetilde{Q}_{(F+1),A} \equiv a_A \tag{207}$$

$$\widetilde{Q}_{A,(F+2)} = \widetilde{Q}_{(F+2),A} \equiv b_A \tag{208}$$

$$\widetilde{Q}_{(F+1),(F+1)} \equiv c \tag{209}$$

$$\widetilde{Q}_{(F+1),(F+2)} = \widetilde{Q}_{(F+2),(F+1)} \equiv d \tag{210}$$

$$\widetilde{Q}_{(F+2),(F+2)} \equiv e \tag{211}$$

Also, let x_a and y_a, $a = 1, \ldots, (F+2)$ be the following $(F+2)$-vectors:

$$x_A \equiv v_A \tag{212}$$

$$x_{F+1} \equiv \mu \tag{213}$$

$$x_{F+2} \equiv \widetilde{\mu} \tag{214}$$

$$y_A = 0 \tag{215}$$

$$y_{F+1} = -1 \tag{216}$$

$$y_{F+2} = -\widetilde{P} \tag{217}$$

Then we have

$$x_a = \sum_{b=1}^{F+2} \widehat{Q}_{ab}\, y_b \tag{218}$$

where $\widehat{Q}_{ab}$ is the matrix inverse to $\widetilde{Q}_{ab}$: $\widehat{Q} \equiv \widetilde{Q}^{-1}$.

Note that (218) solves for v_A, μ and $\widetilde{\mu}$ given η_i, J and J'. On the other hand, (194), (195) and (196) determine η_i, J and J' in terms of v_A, μ and $\widetilde{\mu}$. The entire system is then solved iteratively, where at the initial iteration one takes $J^{(0)} = \{1, \ldots, N\}$, so that $J'^{(0)}$ is empty, and

$$\eta_i^{(0)} = \pm 1, \quad i = 1, \ldots, N \tag{219}$$

While *a priori* the values of $\eta_i^{(0)}$ can be arbitrary, unless $F \ll N$, in some cases one might encounter convergence issues. However, if one chooses

$$\eta_i^{(0)} = \mathrm{sign}(\alpha_i), \quad i = 1, \ldots, N \tag{220}$$

then the iterative procedure generally is expected to converge rather fast. Furthermore, note that the solution is actually exact, *i.e.*, the convergence criteria are given by (from Subsection 5.5 we have that for $\mu \geq 0$ this produces the global optimum)

$$J^{(s+1)} = J^{(s)} \tag{221}$$

$$\forall i \in J^{(s+1)} : \quad \eta_i^{(s+1)} = \eta_i^{(s)} \tag{222}$$

$$\mu^{(s+1)} = \mu^{(s)} \tag{223}$$

$$\widetilde{\mu}^{(s+1)} = \widetilde{\mu}^{(s)} \tag{224}$$

$$\forall A \in \{1, \ldots, F\} : \quad v_A^{(s+1)} = v_A^{(s)} \tag{225}$$

24

where s and $s+1$ label successive iterations.[24] Put differently, the iterative procedure is finite – it converges in a finite number of iterations. Finally, note that w_i for $i \in J$ are given by (193), while $w_i = 0$ for $i \in J'$.

5.3 $\mu = 0$ Case

From the previous subsection we have

$$\mu = \frac{\widetilde{d}\ \widetilde{P} - \widetilde{e}}{\widetilde{c}\ \widetilde{e} - \widetilde{d}^2} \tag{226}$$

$$\widetilde{\mu} = \frac{\widetilde{d} - \widetilde{c}\ \widetilde{P}}{\widetilde{c}\ \widetilde{e} - \widetilde{d}^2} \tag{227}$$

$$0 < \widetilde{R}^2 = 2\ g(w_i, \mu, \widetilde{\mu})|_{\text{optimum}} = -\mu - \widetilde{\mu}\ \widetilde{P} = \frac{\widetilde{c}\ \widetilde{P}^2 - 2\ \widetilde{d}\ \widetilde{P} + \widetilde{e}}{\widetilde{c}\ \widetilde{e} - \widetilde{d}^2} \tag{228}$$

where

$$\widetilde{c} \equiv c - a^T\ Q^{-1}\ a = \sum_{i,j \in J} C_{ij}^{-1}\ \eta_i\ \eta_j > 0 \tag{229}$$

$$\widetilde{d} \equiv d - b^T\ Q^{-1}\ a = \sum_{i,j \in J} C_{ij}^{-1}\ \alpha_i\ \eta_j \tag{230}$$

$$\widetilde{e} \equiv e - b^T\ Q^{-1}\ b = \sum_{i,j \in J} C_{ij}^{-1}\ \alpha_i\ \alpha_j > 0 \tag{231}$$

$$C_{ij}^{-1} = \frac{1}{\xi_i^2}\ \delta_{ij} - \sum_{A,B=1}^{F} \frac{\widetilde{\Omega}_{iA}}{\xi_i^2}\ \frac{\widetilde{\Omega}_{iB}}{\xi_j^2}\ Q_{AB}^{-1} \tag{232}$$

and Q_{AB}^{-1} is the inverse of Q_{AB}.[25] Note that if $\mu \geq 0$, then $\widetilde{\mu} < 0$ assuming $\widetilde{P} = 0$.
   From (226) for $\mu = 0$ we have

$$\widetilde{P} = \widetilde{P}_* \equiv \frac{\widetilde{e}}{\widetilde{d}} > 0 \tag{233}$$

$$\widetilde{\mu} = -\frac{1}{\widetilde{d}} < 0 \tag{234}$$

We will now show that this corresponds to the weights $\widetilde{w}_i$ that maximize the Sharpe ratio: $S \to \max$. (We will denote this maximal value of the Sharpe ratio via $S_{\max}$.)

---

[24] The first two of these criteria are based on discrete quantities and are unaffected by computational (machine) precision effects, while the last three of these criteria are based on continuous quantities and in practice are understood as satisfied within computational (machine) precision. In practice, it suffices to check only one of the three continuous criteria, *e.g.*, the convergence of μ.
   [25] Note that, because C_{ij} has the factor model form, in this case the $N(J) \times N(J)$ matrix D defined in (183) *coincides* with the restriction of the inverse of the $N \times N$ matrix C_{ij}, $i, j \in 1, \dots, N$ to $i, j \in J$.

These weights are given by

$$\widetilde{w}_i = \gamma \sum_{i=1}^{N} C_{ij}^{-1} \, \alpha_j = \frac{\gamma}{\xi_i^2} \left(\alpha_i - \sum_{j=1}^{N} \frac{\alpha_j}{\xi_j^2} \sum_{A,B=1}^{F} \widetilde{\Omega}_{iA} \, \widetilde{\Omega}_{jB} \, Q_{AB}^{-1} \right) \qquad (235)$$

where γ is fixed by (3). So we have ($\widetilde{\eta}_i \equiv \operatorname{sign}(\widetilde{w}_i)$):

$$\widetilde{P}\,\Big|_{S=S_{\max}} = \sum_{i=1}^{N} \alpha_i \, \widetilde{w}_i = \gamma \left(e - \sum_{A,B=1}^{F} Q_{AB}^{-1} \, b_A \, b_B \right) = \gamma \, \widetilde{e} \qquad (236)$$

$$1 = \sum_{i=1}^{N} \widetilde{\eta}_i \, \widetilde{w}_i = \gamma \left(d - \sum_{A,B=1}^{F} Q_{AB}^{-1} \, a_A \, b_B \right) = \gamma \, \widetilde{d} \qquad (237)$$

and we have (233).[26] Note that in this case generically J' is empty and $\widetilde{w}_i$ can only vanish "accidentally" for those alphas such that the expression in the parenthesis in (235) vanishes. In the following we will assume that such "accidental" vanishings do not take place, *i.e.*, let us consider generic alpha configurations. Also, we will denote the values of $\widetilde{c}$, $\widetilde{d}$ and $\widetilde{e}$ at $S = S_{\max}$ via $\widetilde{c}_* > 0$, $\widetilde{d}_* > 0$ and $\widetilde{e}_* > 0$. The second inequality follows from (234).

Next, note that we can focus on the values of $\widetilde{P} \geq \widetilde{P}_*$. Indeed, it makes no sense to consider $\widetilde{P} < \widetilde{P}_*$ as we have a solution with higher P&L and higher Sharpe ratio, namely, the $S = S_{\max}$ solution. Further, let us consider a solution with $\widetilde{P} = (1+\delta)\widetilde{P}_*$, where $0 < \delta \ll 1$. Then $w_i = (1 + \xi_i)\widetilde{w}_i$, where $|\xi_i| \ll 1$ and we have taken into account that none of the $\widetilde{w}_i$ vanishes. We then have $\eta_i = \widetilde{\eta}_i$, and the values of $\widetilde{c}$, $\widetilde{d}$ and $\widetilde{e}$ for w_i are the same as for $\widetilde{w}_i$. So we have

$$\mu = \frac{\widetilde{e}_*}{\widetilde{c}_* \, \widetilde{e}_* - \widetilde{d}_*^2} \, \delta \qquad (238)$$

$$\widetilde{\mu} = -\frac{1 + \widetilde{c}_* \, \mu}{\widetilde{d}_*} \qquad (239)$$

$$\widetilde{R}^2 = \frac{\widetilde{e}_*}{\widetilde{d}_*^2} \, (1 + \widetilde{c}_* \, \mu \, \delta) \qquad (240)$$

Since for $\delta > 0$ the P&L increases, the volatility cannot decrease as $\delta = 0$ corresponds to $S = S_{\max}$. It then follows from (240) that $\mu > 0$ for $0 < \delta \ll 1$ as we have $\widetilde{c}_* > 0$.

Note that $\mu = 0$ uniquely corresponds to the $S = S_{\max}$ solution, so μ cannot become zero at any $\widetilde{P} > \widetilde{P}_*$. Therefore, as we smoothly increase $\widetilde{P}$, the only way μ can become negative is via a discontinuity in $\widetilde{c}$, $\widetilde{d}$ and $\widetilde{e}$, *i.e.*, for some $\widetilde{P} = \widetilde{P}_1 > \widetilde{P}_*$ we have $\mu = \mu_1 > 0$ with one or more vanishing w_i, *i.e.*, $J' \equiv J'_1$ is not empty, while

---

[26] The reason why $\mu = 0$ corresponds to $S \to \max$ can be understood by noting that for $\mu \neq 0$ the term proportional to μ in (164) breaks the scaling property that $g \to \lambda^2 g$ under $w_i \to \lambda w_i$, $\widetilde{\mu} \to \lambda \widetilde{\mu}$ and $\widetilde{P} \to \lambda \widetilde{P}$.

for $\widetilde{P} = (1 + \delta)\widetilde{P}_1$, where $|\delta| \ll 1$, we have empty J' and $\mu = \mu_2(\delta) < 0$. This would imply that as $\delta \to 0$, we would have $w_i \to 0$ for $i \in J'_1$ with non-vanishing $\mu = \mu_2(0) < 0$, which is impossible. (Note that $\widetilde{c}$, $\widetilde{d}$ and $\widetilde{e}$ are unchanged as $\delta \to 0$.) Also, $\mu > 0$ and J' is not empty as $\widetilde{P} \to \max(\alpha_i)$. Finally, note that $\mu < 0$ is compatible with $\widetilde{P} < \widetilde{P}_*$ as no w_i have to vanish for μ to become 0 as $\widetilde{P} \to \widetilde{P}_*-$.

5.4 $\mu < 0$ case

If $\mu < 0$, then we no longer have (194) and (195). Instead, from (178) it follows that J' is empty and $J = \{1, \ldots, N\}$, i.e., there are no vanishing w_i, and we have

$$\mu < 0 \tag{241}$$

$$w_i = \frac{1}{\xi_i^2} \left(\eta_i |\mu| + z_i\right), \quad i = 1, \ldots, N \tag{242}$$

$$z_i \equiv -\left(\widetilde{\mu} \alpha_i + \sum_{A=1}^{F} \widetilde{\Omega}_{iA} v_A\right), \quad i = 1, \ldots, N \tag{243}$$

It then follows that

$$\eta_i = \mathrm{sign}\,(z_i), \quad i \in \widetilde{J} \tag{244}$$

$$\forall i \in \widetilde{J}: \quad |z_i| \geq |\mu| \tag{245}$$

However, η_i can be 1 or -1 for $i \notin \widetilde{J}$, i.e., it is undetermined.

This is because for $\mu < 0$ generically there are expected to be more than one local minima. This is problematic in two ways. First, it makes finding the global minimum more difficult due to the usual instabilities in any iterative procedure, especially when the number of local minima is large. Second, in the presence of a large number of local minima, the system is more prone to the out-of-sample instability as relatively small fluctuations in the off-diagonal elements of C_{ij} can lead to an iterative algorithm jumping between local minima. However, as we discussed above, $\mu < 0$ luckily corresponds to the uninteresting range $\widetilde{P} < \widetilde{P}_*$.

5.5   Conditions for Global Minimum

Above we gave the conditions for the global minimum:

$$\sum_{i,j=1}^{N} C_{ij} (w_i + \epsilon_i) (w_j + \epsilon_j) \geq \sum_{i,j \in J} C_{ij} w_i w_j \tag{246}$$

$$\sum_{i=1}^{N} |w_i + \epsilon_i| = 1 \tag{247}$$

$$\sum_{i=1}^{N} \alpha_i (w_i + \epsilon_i) = \widetilde{P} \tag{248}$$

27

where w_i, $i \in J$ are determined using (168), while $w_i = 0$, $i \in J'$, and ϵ_i are arbitrary subject to (247) and (248). Above we discussed these conditions for arbitrary infinitesimal ϵ_i, which give the conditions for a local minimum. Here we discuss the above conditions for non-infinitesimal ϵ_i. We have

$$\sum_{i,j=1}^{N} C_{ij} \ \epsilon_i \ \epsilon_j + 2 \sum_{i,j \in J} C_{ij} \ w_i \ \epsilon_j + 2 \sum_{i \in J} \sum_{j \in J'} C_{ij} \ w_i \ \epsilon_j \geq 0 \tag{249}$$

$$\sum_{i \in J} \left(|w_i + \epsilon_i| - |w_i| \right) + \sum_{i \in J'} |\epsilon_i| = 0 \tag{250}$$

$$\sum_{i \in J} \alpha_i \ \epsilon_i + \sum_{i \in J'} \alpha_i \ \epsilon_i = 0 \tag{251}$$

which, taking into account (168) and (177), reduce to

$$\sum_{i,j=1}^{N} C_{ij} \ \epsilon_i \ \epsilon_j \geq 2 \ \mu \ \left(\sum_{i \in J} \eta_i \ \epsilon_i + \sum_{i \in J'} |\epsilon_i| \right) \tag{252}$$

Note that for infinitesimal ϵ_i the r.h.s. of (252) vanishes due to (250), while the l.h.s. is positive due to positive-definiteness of C_{ij}. However, for non-infinitesimal ϵ_i we have nontrivial conditions.

Let us define $J_1 \subset J$ and $J_2 \subset J$ as follows:

$$\eta_i \ (w_i + \epsilon_i) \geq 0, \quad i \in J_1 \tag{253}$$
$$\eta_i \ (w_i + \epsilon_i) < 0, \quad i \in J_2 \tag{254}$$

Then, from (250) we have

$$\sum_{i \in J} \eta_i \ \epsilon_i + \sum_{i \in J'} |\epsilon_i| = -2 \sum_{i \in J_2} |w_i + \epsilon_i| \tag{255}$$

and

$$\sum_{i,j=1}^{N} C_{ij} \ \epsilon_i \ \epsilon_j \geq -4 \ \mu \sum_{i \in J_2} |w_i + \epsilon_i| \tag{256}$$

This condition is always satisfied for $\mu \geq 0$, which implies that the solution to the local optimum conditions discussed above is automatically the global optimum. However, for $\mu < 0$ (256) implies that a solution to the local optimum conditions is not guaranteed to be the global optimum. In fact, as we discuss in Subsection 5.4, for $\mu < 0$ we can have multiple local minima.

5.6 General Case

The methods employed in the context of the factor model can be used to tackle the case of a general covariance matrix with the understanding that, as we discuss below, convergence of an iterative procedure in the general case is tricky.

We want to solve (168) subject to (169) and (170) for a general positive-definite covariance matrix C_{ij} (assuming $C_{ii} > 0$). From (168) we have

$$w_i = -\frac{1}{C_{ii}} \; (\mu \; \eta_i + \widetilde{\mu} \; \alpha_i + v_i) \,, \quad i \in J \tag{257}$$

$$v_i \equiv \sum_{j \in J(i)} C_{ij} \; w_j \,, \quad i \in J \tag{258}$$

$$J(i) \equiv J \setminus \{i\} \tag{259}$$

i.e., $J(i)$ is defined as J with the single-element subset $\{i\}$ subtracted, and J (and J') are defined as above. Let

$$Q_{ij} \equiv \frac{C_{ij}}{C_{jj}} \,, \quad i, j \in J \tag{260}$$

$$a_i \equiv \sum_{j \in J(i)} Q_{ij} \; \eta_j = \sum_{j \in J} Q_{ij} \; \eta_j - \eta_i \,, \quad i \in J \tag{261}$$

$$b_i \equiv \sum_{j \in J(i)} Q_{ij} \; \alpha_j = \sum_{j \in J} Q_{ij} \; \alpha_j - \alpha_i \,, \quad i \in J \tag{262}$$

$$\widetilde{a}_i \equiv \frac{\eta_i}{C_{ii}} \,, \quad i \in J \tag{263}$$

$$\widetilde{b}_i \equiv \frac{\alpha_i}{C_{ii}} \,, \quad i \in J \tag{264}$$

$$c \equiv \sum_{i \in J} \frac{1}{C_{ii}} \tag{265}$$

$$d \equiv \sum_{i \in J} \frac{\alpha_i \; \eta_i}{C_{ii}} \tag{266}$$

$$e \equiv \sum_{i \in J} \frac{\alpha_i^2}{C_{ii}} \tag{267}$$

Then, by plugging (257) into (258), (169) and (170), we obtain the following system of $N(J) + 2$ equations for $N(J) + 2$ unknowns v_i, μ and $\widetilde{\mu}$:[27]

$$\sum_{j \in J} Q_{ij} \; v_j + a_i \; \mu + b_i \; \widetilde{\mu} = 0 \,, \quad i \in J \tag{268}$$

$$\sum_{i \in J} \widetilde{a}_i \; v_i + c \; \mu + d \; \widetilde{\mu} = 0 \tag{269}$$

$$\sum_{i \in J} \widetilde{b}_i \; v_i + d \; \mu + e \; \widetilde{\mu} = 0 \tag{270}$$

---

[27] Recall the definition of $N(J) \equiv |J|$.

Let $\widetilde{Q}_{ab}$, $a, b = 1, \ldots, (N(J) + 2)$ be the following $(N(J) + 2) \times (N(J) + 2)$ matrix:

$$\widetilde{Q}_{ij} \equiv Q_{ij}, \quad i, j \in J \tag{271}$$

$$\widetilde{Q}_{i,(N(J)+1)} \equiv a_i, \quad i \in J \tag{272}$$

$$\widetilde{Q}_{(N(J)+1),i} \equiv \widetilde{a}_i, \quad i \in J \tag{273}$$

$$\widetilde{Q}_{i,(N(J)+2)} \equiv b_i, \quad i \in J \tag{274}$$

$$\widetilde{Q}_{(N(J)+2),i} \equiv \widetilde{b}_i, \quad i \in J \tag{275}$$

$$\widetilde{Q}_{(N(J)+1),(N(J)+1)} \equiv c \tag{276}$$

$$\widetilde{Q}_{(N(J)+1),(N(J)+2)} = \widetilde{Q}_{(N(J)+2),(N(J)+1)} \equiv d \tag{277}$$

$$\widetilde{Q}_{(N(J)+2),(N(J)+2)} \equiv e \tag{278}$$

Note that Q_{ab} is not symmetric. Also, let x_a and y_a, $a = 1, \ldots, (N(J) + 2)$ be the following $(N(J) + 2)$-vectors:

$$x_i \equiv v_i, \quad i \in J \tag{279}$$

$$x_{N(J)+1} \equiv \mu \tag{280}$$

$$x_{N(J)+2} \equiv \widetilde{\mu} \tag{281}$$

$$y_i = 0, \quad i \in J \tag{282}$$

$$y_{N(J)+1} = -1 \tag{283}$$

$$y_{N(J)+2} = -\widetilde{P} \tag{284}$$

Then we have

$$x_a = \sum_{b=1}^{N(J)+2} \widehat{Q}_{ab} \, y_b \tag{285}$$

where $\widehat{Q}_{ab}$ is the matrix inverse[28] to $\widetilde{Q}_{ab}$: $\widehat{Q} \equiv \widetilde{Q}^{-1}$.

Next, recalling that $\eta_i \, w_i > 0$, $i \in J$, using (257) and assuming $\mu \geq 0$ we have

$$\mu \geq 0 \tag{286}$$

$$\eta_i = -\text{sign} \left(\widetilde{\mu} \, \alpha_i + v_i \right), \quad i \in J \tag{287}$$

$$\forall i \in J: \quad |\widetilde{\mu} \, \alpha_i + v_i| > \mu \tag{288}$$

$$\forall i \in J': \quad |\widetilde{\mu} \, \alpha_i + v_i| \leq \mu \tag{289}$$

where the last condition follows from (178). The last three conditions determine η_i, J and J' in terms of v_i, μ and $\widetilde{\mu}$, while (285) solves v_i, μ and $\widetilde{\mu}$ given η_i, J and J'. So, one can attempt to solve the entire system iteratively as in the context of the factor model in Subsection 5.2. In this regard, the $N(J)$-vector v_i here is analogous to the F-vector v_A in Subsection 5.2. However, unless $N(J) \ll N$, the analogy with the factor model is only superficial and the convergence is not guaranteed.

---

[28] This inverse exists so long as C_{ij} is invertible, which we assume to be the case.

6 Linear Cost in Weight Optimization

Next, let us include linear cost in the weight optimization problem. Linear costs can be modeled by subtracting a linear penalty from the P&L:

$$P = I \sum_{i=1}^{N} \alpha_i \, w_i - L \, D \tag{290}$$

where L includes all fixed trading costs (SEC fees, exchange fees, broker-dealer fees, *etc.*) and linear slippage.[29] The linear cost assumes no impact, *i.e.*, trading does not affect the stock prices. Also, $D = I \, T$ is the dollar amount traded, and T is the turnover (so the turnover is defined as a percentage). Let τ_i be the turnovers corresponding to individual alphas α_i. If we ignore turnover reduction resulting from combining alphas (or if the internal crossing is switched off), then

$$T = \sum_{i=1}^{N} \tau_i \, |w_i| \tag{291}$$

However, with internal crossing turnover reduction can be substantial and needs to be taken into account. In [58] we proposed a model of turnover reduction, according to which when the number of alphas N is large, the leading approximation (in the $1/N$ expansion) is given by

$$T \approx \rho_* \sum_{i=1}^{N} \tau_i \, |w_i| \tag{292}$$

where $0 < \rho_* \leq 1$ is the turnover reduction coefficient. Let us emphasize that this formula is expected to be a good approximation in the large N limit (so long as the distribution of individual turnovers τ_i is not skewed) regardless of how ρ_* is modeled. In [58] we also proposed a spectral model for estimating ρ_* based on the correlation matrix Ψ_{ij}:

$$\rho_* \approx \frac{\psi^{(1)}}{N\sqrt{N}} \left| \sum_{i=1}^{N} \widetilde{V}_i^{(1)} \right| \tag{293}$$

where $\psi^{(1)}$ is the largest eigenvalue of Ψ_{ij} and $\widetilde{V}_i^{(1)}$ is the corresponding eigenvector normalized such that $\sum_{i=1}^{N} \left(\widetilde{V}_i^{(1)} \right)^2 = 1$.

We then have

$$P = I \sum_{i=1}^{N} \left(\alpha_i \, w_i - L_i \, |w_i| \right) \tag{294}$$

where

$$L_i \equiv L \, \rho_* \, \tau_i > 0 \tag{295}$$

---

[29]Here for the sake of simplicity the linear slippage is assumed to be uniform across all alphas. This is not a critical assumption and can be relaxed, *e.g.*, by modifying the definition of L_i below. In essence, this assumption is made to simplify the discussion of turnover reduction.

We then follow the approach of Section 5, which is now modified as follows.

For the sake of brevity, we will not repeat the discussions of Section 5 whose applicability is evident, and we will use the same notations as in Section 5 and simply modify the corresponding definitions. Furthermore, we will assume the factor model of Subsection 5.1. We need to solve the following problem:

$$\frac{P}{I} = \sum_{i=1}^{N} (\alpha_i \ w_i - L_i \ |w_i|) \equiv \widetilde{P} = \text{fixed} \tag{296}$$

$$\frac{R}{I} = \sqrt{\sum_{i,j=1}^{N} C_{ij} \ w_i \ w_j} \to \min \tag{297}$$

subject to

$$\sum_{i=1}^{N} |w_i| = 1 \tag{298}$$

This problem can be stated as follows:

$$g(w, \mu, \widetilde{\mu}) \equiv \frac{1}{2} \sum_{i,j=1}^{N} C_{ij} \ w_i \ w_j +$$

$$+ \mu \left(\sum_{i=1}^{N} |w_i| - 1 \right) + \widetilde{\mu} \left(\sum_{i=1}^{N} (\alpha_i \ w_i - L_i \ |w_i|) - \widetilde{P} \right) \to \min \tag{299}$$

We now have

$$w_i = -\frac{1}{\xi_i^2} \left(\mu \ \eta_i + \widetilde{\mu} \ \overline{\alpha}_i + \sum_{A=1}^{F} \widetilde{\Omega}_{iA} \ v_A \right), \quad i \in J \tag{300}$$

$$\overline{\alpha}_i \equiv \alpha_i - L_i \ \eta_i, \quad i \in J \tag{301}$$

Recalling that we have $w_i \ \eta_i > 0$, $i \in J$, and assuming that $\mu \geq \widetilde{\mu} \ L_i$, $i \in J$ we get

$$\mu \geq \widetilde{\mu} \ L_i, \quad i = 1, \dots, N \tag{302}$$

$$\eta_i = -\text{sign} \left(\widetilde{\mu} \ \alpha_i + \sum_{A=1}^{F} \widetilde{\Omega}_{iA} \ v_A \right), \quad i \in J \tag{303}$$

$$\forall i \in J : \quad \left| \widetilde{\mu} \ \alpha_i + \sum_{A=1}^{F} \widetilde{\Omega}_{iA} \ v_A \right| > \mu - \widetilde{\mu} \ L_i \tag{304}$$

$$\forall i \in J' : \quad \left| \widetilde{\mu} \ \alpha_i + \sum_{A=1}^{F} \widetilde{\Omega}_{iA} \ v_A \right| \leq \mu - \widetilde{\mu} \ L_i \tag{305}$$

Note that $\mu \geq \widetilde{\mu} \ L_i$, $i \in J'$ follows from (305).[30]

_____

[30] The global minimum conditions in the presence of linear costs are discussed in Subsection 6.1, where we also derive the above conditions for a local minimum.

The rest of the discussion is identical to that in Subsection 5.2 with α_i replaced by $\overline{\alpha}_i$ in all the remaining definitions. We still have (226), (227) and (228), so for $\mu \geq 0$ we still have $\widetilde{\mu} < 0$, which implies that the condition (302) is automatically satisfied for $\mu \geq 0$. *I.e.*, the condition for the existence of the unique minimum with the linear costs ($\mu \geq \widetilde{\mu} \ L_i$) is not as strong as in the absence of costs ($\mu \geq 0$). Furthermore, $\mu = 0$ still corresponds to maximizing the Sharpe ratio $S \to \max$ with linear costs [57]. Let the corresponding $\widetilde{P} \equiv \widetilde{P}_*$. Then we can restrict to $\widetilde{P} \geq \widetilde{P}_*$, for which we have $\mu \geq 0$, which means that the results of Subsection 5.2 apply.

Here the following remark is in order. Because the alphas α_i, $i \in J'$ are no longer traded, we can drop such alphas, if any, recompute ρ_* in (295) using the corresponding correlation matrix $\Psi'_{ij} \equiv \Psi_{ij} \mid_{i,j \in J}$, recompute w_i using such ρ_* and repeat this procedure until the subset J based on which ρ_* is computed is the same as the subset for which $w_i \neq 0$, where w_i are computed based on such ρ_*.[31]

6.1 Global Minimum Conditions with Linear Costs

The discussion in Subsection 5.5 is modified as follows in the presence of linear costs. We have

$$\sum_{i,j=1}^{N} C_{ij} \ (w_i + \epsilon_i) \ (w_j + \epsilon_j) \geq \sum_{i,j \in J} C_{ij} \ w_i \ w_j \tag{306}$$

$$\sum_{i=1}^{N} |w_i + \epsilon_i| = 1 \tag{307}$$

$$\sum_{i=1}^{N} (\alpha_i \ (w_i + \epsilon_i) - L_i \ |w_i + \epsilon_i|) = \widetilde{P} \tag{308}$$

where w_i, $i \in J$ are determined via

$$\sum_{j \in J} C_{ij} \ w_j = -\mu \ \eta_i - \widetilde{\mu} \ (\alpha_i - L_i \ \eta_i), \quad i \in J \tag{309}$$

while $w_i = 0$, $i \in J'$. We have

$$\sum_{i,j=1}^{N} C_{ij} \ \epsilon_i \ \epsilon_j + 2 \sum_{i,j \in J} C_{ij} \ w_i \ \epsilon_j + 2 \sum_{i \in J} \sum_{j \in J'} C_{ij} \ w_i \ \epsilon_j \geq 0 \tag{310}$$

$$\sum_{i \in J} (|w_i + \epsilon_i| - |w_i|) + \sum_{i \in J'} |\epsilon_i| = 0 \tag{311}$$

$$\sum_{i \in J} (\alpha_i \ \epsilon_i - L_i \ (|w_i + \epsilon_i| - |w_i|)) + \sum_{i \in J'} (\alpha_i \ \epsilon_i - L_i \ |\epsilon_i|) = 0 \tag{312}$$

---

[31] When N is large, this procedure is stable and convergent as ρ_* does not change much with N (see [58]).

which, taking into account (309), reduce to

$$
\sum_{i,j=1}^{N} C_{ij} \; \epsilon_i \; \epsilon_j + 2 \sum_{j \in J'} \left(\sum_{i \in J} C_{ij} \; w_i \; \epsilon_j + (\mu - \widetilde{\mu} \; L_j) \;\; |\epsilon_j| + \widetilde{\mu} \; \alpha_j \; \epsilon_j \right) \geq
$$
$$
\geq -4 \sum_{i \in J_2} (\mu - \widetilde{\mu} \; L_i) \;\; |w_i + \epsilon_i| \tag{313}
$$

where, as above, J_2 is defined as follows:

$$
\eta_i \; (w_i + \epsilon_i) < 0, \quad i \in J_2 \subset J \tag{314}
$$

For infinitesimal ϵ_i we have empty J_2 and (313) reduces to

$$
\sum_{j \in J'} \left(\sum_{i \in J} C_{ij} \; w_i \; \epsilon_j + (\mu - \widetilde{\mu} \; L_j) \;\; |\epsilon_j| + \widetilde{\mu} \; \alpha_j \; \epsilon_j \right) \geq 0 \tag{315}
$$

which gives

$$
\forall j \in J' : \quad \left| \sum_{i \in J} C_{ij} \; w_i + \widetilde{\mu} \; \alpha_j \right| \leq \mu - \widetilde{\mu} \; L_j \tag{316}
$$

which in turn gives (305) in the factor model context.

For non-infinitesimal ϵ_i we have the following condition:

$$
\sum_{i,j=1}^{N} C_{ij} \; \epsilon_i \; \epsilon_j \geq -4 \sum_{i \in J_2} (\mu - \widetilde{\mu} \; L_i) \;\; |w_i + \epsilon_i| \tag{317}
$$

which is always satisfied if

$$
\mu \geq \widetilde{\mu} \; L_i, \quad i \in J \tag{318}
$$

In this case the solution to the local optimum conditions discussed above is automatically the global optimum. However, if (318) is not satisfied, then a solution to the local optimum conditions is not guaranteed to be the global optimum. In fact, in this case we can have multiple local minima.

7 Impact in Weight Optimization

Finally, let us discuss the effect of impact, *i.e.*, nonlinear costs, on weight optimization. Generally, introducing nonlinear impact makes the weight optimization problem computationally more challenging and requires introduction of approximation methods.

One way of modeling trading costs is to introduce linear and nonlinear terms:

$$
P = I \sum_{i=1}^{N} \alpha_i \; w_i - L \; D - \frac{1}{n} \; Q \; D^n \tag{319}
$$

where $D = I\,T$ is the dollar amount traded, and T is the turnover. If we model turnover using (292), then we have

$$P = I\,\sum_{i=1}^{N}(\alpha_i\,w_i - L_i\,|w_i|) - \frac{1}{n}\,\widetilde{Q}\,\left(\sum_{i=1}^{N}\tau_i\,|w_i|\right)^{n} \tag{320}$$

where the modulus accounts for the possibility of some w_i being negative, and $\widetilde{Q}$ is defined as follows

$$\widetilde{Q} \equiv Q\,(I\,\rho_*)^n \tag{321}$$

For general fractional n, which would have to be measured empirically, the weight optimization problem would have to be solved numerically.

First, note that if individual turnovers $\tau_i \equiv \tau$ are identical, then the nonlinear cost contribution into P is independent of w_i as we have (3). In this case, it simply shifts $\widetilde{P}$ by a constant and the problem can be solved exactly as in the previous section.[32] If τ_i are not all identical, then we need to solve the following problem:

$$g(w, \mu, \widetilde{\mu}) \equiv \frac{1}{2}\,\sum_{i,j=1}^{N} C_{ij}\,w_i\,w_j + \mu\,\left(\sum_{i=1}^{N}|w_i| - 1\right) +$$

$$+ \widetilde{\mu}\,\left[\sum_{i=1}^{N}(\alpha_i\,w_i - L_i\,|w_i|) - \frac{1}{n}\,\widetilde{Q}'\,\left(\sum_{i=1}^{N}\tau_i\,|w_i|\right)^{n} - \widetilde{P}\right] \tag{322}$$

$$g(w, \mu, \widetilde{\mu}) \to \min \tag{323}$$

where

$$\widetilde{Q}' \equiv \frac{\widetilde{Q}}{I} \tag{324}$$

Here one can use successive iterations to deal with the nonlinear term and various stability issues associated with convergence must be addressed. A simpler approach is, following [57], to note that the key role of the nonlinear term is to model portfolio capacity[33] via its dependence on I, not its detailed structure in terms of individual alphas. In this regard, the following approximation is a reasonable way of simplifying the problem. Let

$$\overline{\tau} \equiv \frac{1}{N}\sum_{i=1}^{N}\tau_i \tag{325}$$

$$\widetilde{\tau}_i \equiv \tau_i - \overline{\tau} \tag{326}$$

---

[32] In fact, in this case the contribution of the linear cost also shifts $\widetilde{P}$ by a constant.

[33] By this we mean the value of the investment level $I = I_*$ for which the P&L $P_{\mathrm{opt}}(I)$ is maximized, where for any given I P&L $P_{\mathrm{opt}}(I)$ is computed for the optimized weights w_i. When only linear cost is present, capacity is unbounded. When nonlinear cost is included, I_* is finite.

If the distribution of $\widetilde{\tau}_i$ has a small standard deviation, then we can use the following approximation:

$$g(w, \mu, \widetilde{\mu}) \approx \frac{1}{2} \sum_{i,j=1}^{N} C_{ij} \ w_i \ w_j + \mu \ \left(\sum_{i=1}^{N} |w_i| - 1 \right) +$$

$$+ \widetilde{\mu} \left[\sum_{i=1}^{N} (\alpha_i \ w_i - L_i \ |w_i|) - \frac{1}{n} \widetilde{Q}' \ \left(\overline{\tau}^n + n \ \overline{\tau}^{n-1} \sum_{i=1}^{N} \widetilde{\tau}_i \ |w_i| \right) - \widetilde{P} \right] \quad (327)$$

The objective function can be rewritten as

$$g(w, \mu, \widetilde{\mu}) \approx \frac{1}{2} \sum_{i,j=1}^{N} C_{ij} \ w_i \ w_j + \mu \ \left(\sum_{i=1}^{N} |w_i| - 1 \right) +$$

$$+ \widetilde{\mu} \ \left(\sum_{i=1}^{N} \left(\alpha_i \ w_i - \widetilde{L}_i \ |w_i| \right) - \widetilde{P}' \right) \quad (328)$$

where

$$\widetilde{L}_i \equiv L_i + \widetilde{Q}' \ \overline{\tau}^{n-1} \ \tau_i = L_i + Q \ \rho_*^n \ I^{n-1} \ \overline{\tau}^{n-1} \ \tau_i \quad (329)$$

$$\widetilde{P}' \equiv \widetilde{P} + \frac{1}{n} \widetilde{Q}' \ \overline{\tau}^n \quad (330)$$

I.e., in this approximation the effect of the nonlinear term reduces to increasing the linear slippage and shifting $\widetilde{P}$, and this problem we can solve as in the previous section. Note, however, that the "effective" linear cost $\widetilde{L}_i$ now depends on the investment level I via (329), which now controls capacity. Thus, for I such that

$$\widetilde{L}_i \geq |\alpha_i| \quad (331)$$

the P&L cannot be positive.[34]

8   Comments

Above, among other things, we discussed optimization where one maximizes P&L subject to a lower bound on the Sharpe ratio. Theoretically this is a perfectly sound optimization problem. However, in practice it needs to be amended for the following reason. Realized return and expected return generally are vastly different. Because of this, the realized Sharpe ratio and the expected Sharpe ratio generally are also vastly different. For this reason, the condition $S \geq S_{\min}$ is impractical in terms of

---

[34] The restriction on I is even more severe, because this condition tells us that $\widetilde{P}'$ cannot be positive, and for $\widetilde{P}$ to be positive, $\widetilde{L}_i$ would have to be even lower. At the end one is interested in determining capacity I_*, for which the optimized P&L is maximized.

the realized Sharpe ratio as there is no natural way of setting realized $S_{\min}$, *i.e.*, the lower bound on the realized Sharpe ratio. However, unlike the Sharpe ratio, portfolio risk R is more stable and it makes sense to replace the condition $S \geq S_{\min}$ by $R \leq R_{\max}$. *I.e.*, one now has the following optimization problem: $P \to \max$, $R \leq R_{\max}$, where the condition on R can now be treated as a condition on expected risk. Happily, the actual solution to the latter optimization problem is the same as that discussed in Sections 5, 6 and 7, where now one specifies an upper bound on risk as opposed to a lower bound on the Sharpe ratio. The same applies to our discussion in Subsection 3.3.

One of our main points above is that using factor models for alpha streams substantially reduces the number of iterations in the optimization problem and renders it practically tractable. Using a constructed covariance matrix as opposed to the one computed based on the alpha time series also has the advantage that – assuming the factor covariance matrix is not computed using the very same time series – the factor model covariance matrix is expected to be more stable out-of-sample. Also, the number of nonzero eigenvalues in the computed covariance matrix is limited to $M \ll N$ (where $M + 1$ is the number of observations – see Section 2), *i.e.*, the number of risk factors (which are essentially based on principal components of the computed covariance matrix) is limited to M, whereas the constructed factor model covariance matrix can have many more risk factors as long as the factor covariance matrix Φ_{AB} is constructed based on stable data. Factor models for alpha streams is discussed in more detail in [60].

References

[1] T. Schneeweis, R. Spurgin, and D. McCarthy, "Survivor Bias in Commodity Trading Advisor Performance", J. Futures Markets, 1996, 16(7), 757-772.

[2] C. Ackerman, R. McEnally and D. Revenscraft, "The Performance of Hedge Funds: Risk, Return and Incentives", Journal of Finance, 1999, 54(3), 833-874.

[3] S.J. Brown, W. Goetzmann and R.G. Ibbotson, "Offshore Hedge Funds: Survival and Performance, 1989-1995", Journal of Business, 1999, 72(1), 91-117.

[4] F.R. Edwards and J. Liew, "Managed Commodity Funds", Journal of Futures Markets, 1999, 19(4), 377-411.

[5] F.R. Edwards and J. Liew, "Hedge Funds versus Managed Futures as Asset Classes", Journal of Derivatives, 1999, 6(4), 45-64.

[6] W. Fung and D. Hsieh, "A Primer on Hedge Funds", Journal of Empirical Finance, 1999, 6(3), 309-331.

[7] B. Liang, "On the Performance of Hedge Funds", Financial Analysts Journal, 1999, 55(4), 72-85.

[8] V. Agarwal and N.Y. Naik, "On Taking the "Alternative" Route: The Risks, Rewards, and Performance Persistence of Hedge Funds", Journal of Alternative Investments, 2000, 2(4), 6-23.

[9] V. Agarwal and N.Y. Naik, "Multi-Period Performance Persistence Analysis of Hedge Funds Source", Journal of Financial and Quantitative Analysis, 2000, 35(3), 327-342.

[10] W. Fung and D. Hsieh, "Performance Characteristics of Hedge Funds and Commodity Funds: Natural vs. Spurious Biases", Journal of Financial and Quantitative Analysis, 2000, 35(3), 291-307.

[11] B. Liang, "Hedge Funds: The Living and the Dead", Journal of Financial and Quantitative Analysis, 2000, 35(3), 309-326.

[12] C.S. Asness, R.J. Krail, and J.M. Liew, "Do Hedge Funds Hedge?", Journal of Portfolio Management, 2001, 28(1), 6-19.

[13] F.R. Edwards and M.O. Caglayan, "Hedge Fund and Commodity Fund Investments in Bull and Bear Markets", Journal of Portfolio Management, 2001, 27(4), 97-108.

[14] W. Fung and D. Hsieh, "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers", Review of Financial Studies, 2001, 14(2), 313-341.

[15] B. Liang, "Hedge Fund Performance: 1990-1999", Financial Analysts Journal, 2001, 57(1), 11-18.

[16] A.W. Lo, "Risk Management For Hedge Funds: Introduction and Overview", Financial Analysis Journal, 2001, 57(6), 16-33.

[17] C. Brooks and H.M. Kat, "The Statistical Properties of Hedge Fund Index Returns and Their Implications for Investors", Journal of Alternative Investments, 2002, 5(2), 26-44.

[18] D.-L. Kao, "Battle for Alphas: Hedge Funds versus Long-Only Portfolios", Financial Analysts Journal, 2002, 58(2), 16-36.

[19] G. Amin and H. Kat, "Stocks, Bonds and Hedge Funds: Not a Free Lunch!", Journal of Portfolio Management, 2003, 29(4), 113-120.

[20] N. Chan, M. Getmansky, S.M. Haas and A.W. Lo, "Systemic Risk and Hedge Funds", published in: Carey, M. and Stulz, R.M., eds., "The Risks of Financial Institutions" (University of Chicago Press, 2006), Chapter 6, 235-338.

[21] Z. Kakushadze and J.K.-S. Liew, "Is It Possible to OD on Alpha?", The Journal of Alternative Investments (forthcoming); http://ssrn.com/abstract=2419415.

[22] H. Markowitz, "Portfolio selection", Journal of Finance, 1952, 7(1), 77-91.

[23] A. Charnes and W.W. Cooper, "Programming with linear fractional functionals", Naval Research Logistics Quarterly, 1962, 9(3-4), 181-186.

[24] W.F. Sharpe, "Mutual fund performance", Journal of Business, 1966, 39(1), 119-138.

[25] R.C. Merton, "Lifetime portfolio selection under uncertainty: the continuous time case", The Review of Economics and Statistics, 1969, 51(3), 247-257.

[26] S. Schaible, "Parameter-free convex equivalent and dual programs of fractional programming problems", Zeitschrift für Operations Research, 1974, 18(5), 187-196.

[27] M. Magill and G. Constantinides, "Portfolio selection with transactions costs", J. Econom. Theory, 1976, 13(2), 245-263.

[28] A.F. Perold, "Large-scale portfolio optimization", Management Science, 1984, 30(10), 1143-1160.

[29] M. Davis and A. Norman, "Portfolio selection with transaction costs", Math. Oper. Res., 1990, 15(4), 676-713.

[30] B. Dumas and E. Luciano, "An exact solution to a dynamic portfolio choice problem under transaction costs", The Journal of Finance, 1991, 46(2), 577-595.

[31] C. J. Adcock and N. Meade "A simple algorithm to incorporate transactions costs in quadratic optimization", European Journal of Operational Research, 1994, 79(1), 85-94.

[32] S. Shreve and H.M. Soner, "Optimal investment and consumption with transaction costs", Ann. Appl. Probab., 1994, 4(3), 609-692.

[33] D. Bienstock, "Computational study of a family of mixed-integer quadratic programming problems", Mathematical Programming, 1996, 74(2), 121-140.

[34] J. Cvitanić and I. Karatzas, "Hedging and portfolio optimization under transaction costs: a martingale approach", Math. Finance, 1996, 6(2), 133-165.

[35] A. Yoshimoto, "The mean-variance approach to portfolio optimization subject to transaction costs", J. Operations Research Soc. of Japan, 1996, 39(1), 99-117.

[36] C. Atkinson, S.R. Pliska and P. Wilmott, "Portfolio management with transaction costs", Proc. Roy. Soc. London Ser. A, 1997, 453(1958), 551-562.

[37] D. Bertsimas, C. Darnell and R. Soucy, "Portfolio construction through mixed-integer programming at Grantham, Mayo, Van Otterloo and Company", Interfaces, 1999, 29(1), 49-66.

[38] A. Cadenillas and S. R. Pliska, "Optimal trading of a security when there are taxes and transaction costs", Finance and Stochastics, 1999, 3(2), 137-165.

[39] T.-J. Chang, N. Meade, J.E. Beasley and Y.M. Sharaiha, "Heuristics for cardinality constrained portfolio optimisation", Computers and Operations Research, 2000, 27(13), 1271-1302.

[40] H. Kellerer, R. Mansini and M.G. Speranza, "Selecting portfolios with fixed costs and minimum transaction lots", Annals of Operations Research, 2000, 99(1-4), 287-304.

[41] R.T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk", Journal of Risk, 2000, 2(3), 21-41.

[42] J. Gondzio and R. Kouwenberg, "High-performance computing for asset-liability management", Operations Research, 2001, 49(6), 879-891.

[43] H. Konno and A. Wijayanayake, "Portfolio optimization problem under concave transaction costs and minimal transaction unit constraints", Mathematical Programming, 2001, 89(2), 233-250.

[44] S. Mokkhavesa and C. Atkinson, "Perturbation solution of optimal portfolio theory with transaction costs for any utility function", IMA J. Manag. Math., 2002, 13(2), 131-151.

[45] O.L.V. Costa and A.C. Paiva, "Robust portfolio selection using linear-matrix inequalities", Journal of Economic Dynamics and Control, 2002, 26(6), 889-909.

[46] F. Alizadeh and D. Goldfarb, "Second-order cone programming", Mathematical Programming, 2003, 95(1), 3-51.

[47] M.J. Best and J. Hlouskova, "Portfolio selection and transactions costs", Computational Optimization and Applications, 2003, 24(1), 95-116.

[48] K. Janeček and S. Shreve, "Asymptotic analysis for optimal investment and consumption with transaction costs", Finance Stoch., 2004, 8(2), 181-206.

[49] M.S. Lobo, M. Fazel and S. Boyd, "Portfolio optimization with linear and fixed transaction costs", Annals of Operations Research, 2007, 152(1), 341-365.

[50] R. Zagst and D. Kalin, "Portfolio optimization under liquidity costs", International Journal of Pure and Applied Mathematics, 2007, 39(2), 217-233.

[51] M. Potaptchik, L. Tunçel and H. Wolkowicz, "Large scale portfolio optimization with piecewise linear transaction costs", Optimization Methods and Software, 2008, 23(6), 929-952.

[52] E. Moro, J. Vicente, L.G. Moyano, A. Gerig, J.D. Farmer, G. Vaglica, F. Lillo and R.N. Mantegna, "Market impact and trading profile of hidden orders in stock markets", Physical Review E, 2009, 80, 066102.

[53] J. Goodman and D.N. Ostrov, "Balancing small transaction costs with loss of optimal allocation in dynamic stock trading strategies", SIAM J. Appl. Math., 2010, 70(6), 1977-1998.

[54] M. Bichuch, "Asymptotic analysis for optimal investment in finite time with transaction costs", SIAM J. Financial Math., 2012, 3(1), 433-458.

[55] J.E. Mitchell and S. Braun, "Rebalancing an investment portfolio in the presence of convex transaction costs, including market impact costs", Optimization Methods and Software, 2013, 28(3), 523-542.

[56] H. Soner and N. Touzi, "Homogenization and asymptotics for small transaction costs", SIAM Journal on Control and Optimization, 2013, 51(4), 2893-2921.

[57] Z. Kakushadze, "Combining Alpha Streams with Costs", The Journal of Risk, 2015, 17(3), 57-78; http://ssrn.com/abstract=2438687; arXiv:1405.4716

[58] Z. Kakushadze, "Spectral Model of Turnover Reduction", SSRN Working Paper, http://ssrn.com/abstract=2427049; arXiv:1404.5050.

[59] R. Rebonato and P. Jäckel, "The most general methodology to create a valid correlation matrix for risk management and option pricing purposes" (1999), http://ssrn.com/abstract=1969689 (December 7, 2011).

[60] Z. Kakushadze, "Factor Models for Alpha Streams", The Journal of Investment Strategies, 2014, 4(1), 83-109; http://ssrn.com/abstract=2449927; arXiv:1406.3396 [q-fin.PM].