Understanding the Long-Only Minimum Variance Portfolio

– DRAFT –

Alec Kercheval and Ololade Sowunmi
Florida State University
osowunmi@fsu.edu akercheval@fsu.edu
Nicholas Gunther
nlgunther@gmail.com

First version: 10 Dec 2024; This version: 15 June 2025

Abstract

For a covariance matrix coming from a factor model of returns, we investigate the relationship between the long-only global minimum variance portfolio and the asset exposures to the factors. In the case of a 1-factor model, we provide a rigorous and explicit description of the long-only solution in terms of the parameters of the covariance matrix. For $q > 1$ factors, we provide a description of the long-only portfolio in geometric terms. The results are illustrated with empirical daily returns of US stocks.

1 Long-only minimum variance

1.1 Introduction There is a long-standing interest in equity portfolios optimized to have the lowest possible variance. The optimal such portfolio depends on the covariances between pairs of assets, and on the particular constraints of interest.

If there are p assets available for investment, we denote by $w = (w_1, \ldots, w_p)^\top \in \mathbf{R}^p$ the p-dimensional vector of asset weights defining the portfolio. The global minimum variance portfolio w^{LS} denotes the long-short portfolio solving the simplest problem

$$\begin{aligned} &\min_{w \in \mathbf{R}^p} w^\top \Sigma w \\ &w^\top \mathbf{1}_p = 1, \end{aligned} \tag{1}$$

where Σ denotes the positive definite covariance matrix of asset returns; $\mathbf{1}_p$ denotes the vector of dimension p whose every entry is 1; and $w^\top \mathbf{1}_p = 1$ is the full investment condition setting the sum of the weights equal to 1. For the long-short portfolio, some of the weights may be negative.

Our focus is the long-only minimum variance (LOMV) problem

$$\begin{aligned} &\min_{w \in \mathbf{R}^p} w^\top \Sigma w \\ &w^\top \mathbf{1}_p = 1 \\ &w_i \geq 0 \text{ for all } i = 1, 2, ..., p. \end{aligned} \tag{2}$$

The long-only constraints $w_i \geq 0$ are often required for real investment portfolios due to the complications and costs of short positions.

The solution $w = w^L$ of (2) represents the long-only fully invested global minimum risk portfolio, and can be contrasted with the solution w^{LS} of the long-short problem. The portfolio w^{LS} solving problem (1) is given by the simple formula

$$w^{LS} = \frac{\Sigma^{-1}\mathbf{1}_p}{\mathbf{1}_p^\top \Sigma^{-1} \mathbf{1}_p}. \tag{3}$$

The long-only problem (2) is less straightforward.

As we show in this article, the main difficulty in problem (2) is determining which are the active (positive weight) assets in the optimal portfolio, or, equivalently, which are the assets for which the long-only constraints are binding. Denote by $K = \{i \leq p : w_i^L > 0\}$ the set of active assets in the long-only optimal portfolio, and let $k \leq p$ denote the number of elements of K . Once we have determined K, Theorem 1 solves the problem: if we denote by Σ^K the $k \times k$ matrix obtained from Σ by deleting all the rows and columns not in K, then the k positive weights of w^L are given by the corresponding entries of the k-dimensional vector

$$w^K = \frac{(\Sigma^K)^{-1}\mathbf{1}_k}{\mathbf{1}_k^\top (\Sigma^K)^{-1} \mathbf{1}_k}. \tag{4}$$

In short, *the active long-only minimum risk portfolio holdings are those of the long-short minimum risk portfolio corresponding to a reduced set of available assets.* This is intuitively reasonable[1] because the long-only constraints are not binding on the positive holdings of w^L.

This leaves the problem of determining K, the set of active assets of w^L, which is normally much smaller than the set of positive-weight assets in the long-short portfolio w^{LS} (e.g. Figure (2)). In this article we analyze that problem when the covariance matrix comes from a factor model. In the case of a single-factor model, we provide an essentially explicit description of K below in Theorem 2. There, one factor explains covariance between assets, and the covariance matrix takes the form

$$\Sigma = \sigma^2 \beta\beta^\top + \Delta, \tag{5}$$

where $\sigma^2 > 0$ is the factor return variance, β is a p-vector of exposures to the factor, and Δ is a diagonal matrix of specific variances.

In the case of a multiple factor model with $q > 1$ factors, the $p \times p$ covariance matrix takes the form

$$\Sigma = B\Omega B^\top + \Delta, \tag{6}$$

[1]Geometrically, if, for example, the active LOMV assets are the first k in the list, then the corresponding k-vector minimizes k-dimensional variance in the interior of the positive k-orthant.

where B is a $p \times q$ matrix whose columns are the asset exposures to each of q factors, and Ω is a $q \times q$ invertible matrix of factor variances, diagonal in the case when the columns of B are principal components. When $q > 1$, we know of no direct way to compute the set K of active assets that is similar to Theorem 2. However, Theorem 3 gives us a necessary condition satisfied by K, as follows. Let $B_i \in \mathbf{R}^q$ denote the ith row of B. Then there is a $(q-1)$-dimensional hyperplane H in $\mathbf{R}^q$ such that the elements i of K are all those such that B_i lies on the same side of H as the origin.

The single factor case was previously studied in Clarke et al. [2011], where they assume the single factor is the market return, and provide an implicit solution to the long-only problem that is equivalent to the one here when our vector beta of exposures to the single factor is taken to be the market beta. Their article inspired our work, and we discuss the relationship between their results and ours further below.

1.2 Main Results We state the results outlined above in more detail in this section. The proofs of the following theorems appear in Section 3.

Assumption for Theorem 1. Suppose that the covariance matrix Σ of returns for a universe of p assets is an arbitrary $p \times p$ symmetric positive definite matrix.

Theorem 1. *Denote by w^L the solution of problem (2) and let K denote the set of active assets in w^L:*

$$K = \{i \leq p : w_i^L > 0\}, \tag{7}$$

and $k = |K| \leq p$, the number of active assets. Let $\Sigma^{K,0}$ be the modified matrix obtained from Σ by setting to zero the rows and columns not belonging to K.

Then

$$w^L = \frac{(\Sigma^{K,0})^+ \mathbf{1}_p}{\mathbf{1}_p^\top (\Sigma^{K,0})^+ \mathbf{1}_p} \tag{8}$$

where $^+$ denotes the Moore-Penrose inverse.[2]

Equivalently, if we denote by w^K the k-vector obtained by deleting all the zero entries of the p-vector w^L, and Σ^K denotes the $k \times k$ matrix obtained from Σ by deleting the rows and columns not belonging to K, then

$$w^K = \frac{(\Sigma^K)^{-1} \mathbf{1}_k}{\mathbf{1}_k^\top (\Sigma^K)^{-1} \mathbf{1}_k}. \tag{9}$$

This is the unique solution of the k-dimensional long-short problem (1) with Σ replaced by Σ^K, and w^L can be recovered from w^K by adding back zero entries for the deleted assets.

Theorem 1 allows us to determine the long-only minimum risk portfolio as soon as we know the set K of active assets. It remains only to determine K, and for the single-index case $q = 1$ this can be accomplished by an explicit method described next in Theorem 2.

[2]For a symmetric matrix S with singular value decomposition $S = UDU^\top$ for orthogonal U and diagonal D, the Moore-Penrose inverse is defined by $S^+ = UD^+U^\top$, where the diagonal matrix D^+ is obtained by replacing each nonzero element by its inverse.

Assumptions for Theorem 2. Assume now a single-factor returns model which gives the covariance matrix Σ the form

$$\Sigma = \sigma^2 \beta\beta^\top + \Delta, \tag{10}$$

where $\sigma^2 > 0$ is the factor return variance, β is a p-vector of factor exposures, and $\Delta = diag(\delta_1^2, \delta_2^2, ..., \delta_p^2)$ is a diagonal matrix of non-zero idiosyncratic asset return variances $\delta_i^2 > 0$.

We permit the entries β_i of β to be positive, negative, or zero, but assume the generic condition

$$\sum_{i=1}^{p} \frac{\beta_i}{\delta_i^2} \neq 0. \tag{11}$$

Notice that the vector β may be replaced by $-\beta$ without changing Σ, so without loss of generality we choose the sign so that

$$\sum_{i=1}^{p} \frac{\beta_i}{\delta_i^2} > 0, \tag{12}$$

as would be the case if all the betas were positive.

In addition, by re-ordering the assets if necessary, for convenience we further assume without loss of generality that the betas are arranged in increasing order:

$$\beta_1 \leq \beta_2 \leq \cdots \leq \beta_p. \tag{13}$$

With these assumptions, our main result is

Theorem 2. ***(Explicit Solution to the long-only constrained problem under a 1-factor model)*** *Assume the betas β_i are arranged in increasing order.*

1. *Let $R_1 = \frac{1}{\sigma^2}$, and for $2 \leq i \leq p$, let*

$$R_i = \frac{1}{\sigma^2} + \sum_{j=1}^{i-1} \frac{\beta_j}{\delta_j^2}(\beta_j - \beta_i).$$

Then there exists $s \leq p$ such that the initially positive sequence $\{R_i\}$ is monotonically increasing until $i = s$, then monotonically decreasing for $i > s$. That is,

$$0 < R_1 \leq R_2 \leq \cdots \leq R_s \geq R_{s+1} \geq \cdots \geq R_p. \tag{14}$$

In particular, the sequence $\{R_i\}$ crosses zero at most once.

2. *Let w^L denote the solution of problem (2), $K = \{i \leq p : w_i^L > 0\}$, and $k = |K|$.*

Then

$$k = \max\left\{i \leq p : R_i > 0\right\} \text{ and } K = \{1, 2, \ldots, k\}. \tag{15}$$

Applying Theorem 1, we may conclude the following. Let $\Sigma^{K,0}$ be the block-diagonal matrix obtained from Σ by replacing all but the first k rows and columns with zeros. Then the solution w^L is given by

$$w^L = \frac{(\Sigma^{K,0})^+ \mathbf{1}_p}{\mathbf{1}_p^\top (\Sigma^{K,0})^+ \mathbf{1}_p} \tag{16}$$

where $+$ denotes the Moore-Penrose inverse.

For an equivalent formulation in terms of ordinary matrix inverse, let Σ^K be the $k \times k$ submatrix of Σ consisting of the first k rows and columns. Denote by w^K,

$$w^K = \frac{(\Sigma^K)^{-1}\mathbf{1}_k}{\mathbf{1}_k^\top(\Sigma^K)^{-1}\mathbf{1}_k}, \tag{17}$$

the long-short fully invested minimum variance solution for the first k assets.

Then the solution $w^L = (w_1^L, \ldots, w_p^L)^\top$ of (2) is given by

$$\begin{aligned} w_i^L &= w_i^K \text{ for } i = 1, ..., k \\ w_i^L &= 0 \text{ for } i = k+1, \ldots, p. \end{aligned} \tag{18}$$

We note that $k = p$ if the long-short fully invested minimum variance portfolio happens to be long-only already. Otherwise, $k < p$, $R_p < 0$, and the sequence $\{R_i\}$ crosses zero exactly once. In this situation, k has the property

$$R_k > 0 \text{ and } R_{k+1} = R_k + (\beta_k - \beta_{k+1}) \sum_{j=1}^{k} \frac{\beta_j}{\delta_j^2} \leq 0. \tag{19}$$

This means that the threshold index k and the solution w are not influenced by the values of δ_j for $j > k$, nor by the values of β_j for $\beta_j > \beta_{k+1}$.

The proof of Theorem 2 also establishes, via Lemma 4 below, the following

Corollary 1. *Under the assumptions above, if w is the solution of problem* (2)*, then*

$$w_i > 0 \text{ if and only if } \beta_i < \frac{\frac{1}{\sigma^2} + \sum_{j=1}^{k} \frac{\beta_j^2}{\delta_j^2}}{\sum_{j=1}^{k} \frac{\beta_j}{\delta_j^2}}. \tag{20}$$

The solution of problem (2) in semi-explicit form was previously described by R. Clarke, H. de Silva, and S. Thorley in Clarke et al. [2011]. They give the following condition, which is closely related to Corollary 1:

$$w_i > 0 \text{ if and only if } \beta_i < \tau, \tag{21}$$

where τ is the solution of the equation

$$\tau = \frac{\frac{1}{\sigma^2} + \sum_{\beta_i < \tau} \frac{\beta_j^2}{\delta_j^2}}{\sum_{\beta_i < \tau} \frac{\beta_j}{\delta_j^2}}. \tag{22}$$

The solution described in Clarke et al. [2011] is equivalent to (17) and (20). Both solutions require the assumption (12), as a simple argument shows: In the absence of any hypotheses about the signs of the betas, replacing β by $-\beta$ leaves the covariance matrix Σ, and hence the solution w, unchanged. But in that case the long positions of w would correspond to betas above a threshold, not below it, contradicting the threshold condition.

The need for assumption (12) is easily overlooked because it likely always holds for betas actually observed in the market. In our one-factor world, if the betas are defined relative to a benchmark portfolio w_B that belongs to our investable universe of p assets, i.e.

$$w_B \in \mathbf{R}^p \text{ and } \beta = \frac{\Sigma w_B}{\sigma_B^2}, \tag{23}$$

then a calculation similar to the ones in Section 3 shows that (12) holds if and only if the benchmark w_B is net long, $w_B^\top \mathbf{1}_p > 0$. This will be the case for any reasonable benchmark.

The closed form solution (17) depends on first determining k from (15). From the monotonicity properties of $\{R_i\}$, this may be quickly accomplished, for example, by the bisection method in $O(\log p)$ steps.

The multifactor case. When there are $q > 1$ factors, there is no simple way to order the betas or immediately use them directly to identify the active assets in the long-only portfolio. The next theorem tells us that we do know something about them: their corresponding betas are exactly those that lie in a particular half-space in $\mathbf{R}^q$.

Assumptions for Theorem 3. We assume asset returns follow a q-factor model, so that the covariance matrix of asset returns takes the form of the $p \times p$ matrix

$$\Sigma = B\Omega B^\top + \Delta, \tag{24}$$

where B is a $p \times q$ matrix whose columns are the asset exposures to the q factors, Ω is a $q \times q$ diagonal matrix of positive factor variances, and Δ is a $p \times p$ diagonal matrix of positive asset specific variances.

As before, for any subset K of the indices $\{1, 2, \ldots, p\}$, if $k = |K|$, we denote by $\Sigma^{K,0}$ the $p \times p$ matrix obtained from Σ by setting all the rows and columns not in K to zero, Σ^K the $k \times k$ principal submatrix obtained by deleting the rows and columns not in K, and B^K the $k \times q$ matrix obtained by deleting the rows of B not in K.

Theorem 3 (Hyperplane separation for q-factor models). *Suppose that w^L is the solution of problem (2) for the covariance matrix (24). Let*

$$K = \{i \leq p : w_i^L > 0\} \tag{25}$$

and $k = |K|$ the cardinality of K.

Define the q-dimensional column vector h^K by

$$h^K = \Omega (B^K)^\top (\Sigma^K)^{-1} \mathbf{1}_k = \Omega B^\top (\Sigma^{K,0})^+ \mathbf{1}_p. \tag{26}$$

Then

$$w_i^L > 0 \text{ if and only if } B_i h^K < 1, \tag{27}$$

where B_i the q-dimensional ith row of B.

Geometrically, the condition of Theorem 3 can be described in terms of the hyperplane H of $\mathbf{R}^q$ defined by

$$H = \{x \in \mathbf{R}^q : x^\top h^K = 1\}. \tag{28}$$

It says that the assets included in the long-only optimal portfolio are those whose factor exposure vectors B_i in $\mathbf{R}^q$ lie on the same side of H as the origin. See Figure 5 for an illustration when $q = 2$.

The next corollary gives alternative hyperplane formulation that is sometimes useful.

Corollary 2. *With the same assumptions and notation of Theorem 3, let*

$$h^L = \left(\Omega^{-1} + (B^K)^\top (\Delta^K)^{-1} B^K\right)^{-1} (B^K)^\top (\Delta^K)^{-1} \mathbf{1}_k \tag{29}$$

$$= \left(\Omega^{-1} + B^\top (\Delta^{K,0})^+ B\right)^{-1} B^\top (\Delta^{K,0})^+ \mathbf{1}_p. \tag{30}$$

Then $B^K h^L = B^K h^K$, and hence $w_i^L > 0$ if and only if $B_i h^L < 1$.

We note that in the typical case where B^K has full rank $q < k$, $B^K h^L = B^K h^K$ implies $h^L = h^K$, so the the formulations of Theorem 3 and Corollary 2 are equivalent.

Unlike Theorem 2, Theorem 3 does not provide a direct computational method for determining w^L. Instead, it provides a necessary condition satisfied by the active assets in terms of their q-vectors B_i, $i = 1, \ldots, p$, as a generalization of the 1-factor setting: the included B_i are the ones below a certain threshold, where the threshold in q dimensions is determined by a $(q-1)$-dimensional hyperplane H.

In typical applications, such as in the illustration below in section 2.3, p is much greater than the number n of samples. In this case, when the factor exposures are bounded and have a positive variance across assets, we expect that $B^K(\Delta^K)^{-1}B^K$ will dominate Ω^{-1} by a factor proportional to p. In the limiting case where Ω^{-1} is set to zero, h^L in equation (29) can be viewed as a vector of coefficients of a weighted least squares regression of $\mathbf{1}$ onto the columns of B^K with weight matrix Δ^{-1}.[3]

In the one-factor case $q = 1$, the condition of Theorem 3 reduces to (20): in this case, B is a single column vector corresponding to β in the single index model, and Ω is a scalar σ^2. A computation using the Woodbury identity for the inverse of Σ^K,

$$(\Sigma^K)^{-1} = (\Delta^K)^{-1}\{I - B^K \left[\frac{1}{\sigma^2} + (B^K)^\top (\Delta^K)^{-1} B^K\right]^{-1} (B^K)^\top (\Delta^K)^{-1}\}, \tag{31}$$

shows that the scalar $h^K = \sigma^2 (B^K)^\top (\Sigma^K)^{-1} \mathbf{1}_k$ is given by

$$h^K = \frac{\sum_{j \in K} \frac{B_j}{\delta_j^2}}{\frac{1}{\sigma^2} + \sum_{j \in K} \frac{B_j^2}{\delta_j^2}} \tag{32}$$

and the condition $B_i h^K < 1$ is equivalent to (20).

2 Numerical Examples

2.1 Single factor exposure estimation Given observed excess returns of p assets over n observation periods, we may form the $p \times n$ data matrix Y and the resulting sample covariance matrix $S = \frac{1}{n} Y Y^\top$.

[3] We thank Lisa Goldberg for this observation.

In the likely event that $p > n$, the matrix S will be rank deficient, so we need a model for estimating an invertible covariance matrix for use in portfolio optimization.

A standard approach is to use factor models to accomplish this. In this section we consider a one-factor model of returns,

$$r = \beta f + \epsilon \tag{33}$$

where β is an unknown vector of factor exposures, f is a random variable with mean zero and variance σ^2 representing the factor return, and ϵ is a random vector whose entries are mutually independent and independent of f, with mean zero and covariance Δ. The columns of Y are then assumed to be n independent realizations of r. The covariance matrix of r is given by (10):

$$\Sigma = \sigma^2 \beta\beta^\top + \Delta.$$

Setting $\zeta^2 = |\beta|^2\sigma^2$ and $b = \beta/|\beta|$, the one-factor covariance model may be written

$$\Sigma = \zeta^2 bb^\top + \Delta, \tag{34}$$

where recall $\Delta = diag(\delta_1^2, \ldots, \delta_p^2)$. Only Y and the corresponding sample covariance matrix S are observed. Estimating Σ requires estimating the scalar ζ^2, the unit vector b, and the vector $(\delta_1^2, \ldots, \delta_p^2)$ of idiosyncratic variances.

For the purpose of computing a minimum variance portfolio, we will consider three different data-driven estimators of (34): $\Sigma^{\text{JSE}}, \Sigma^{\text{MJSE}}$, and Σ^{MS}, defined as follows.

The estimator Σ^{JSE} is a Bayesian-style James-Stein shrinkage estimator with the advantage that the shrinkage takes place entirely inside the class of single-factor models:

$$\Sigma^{\text{JSE}} = \eta^2 h^{\text{JSE}} (h^{\text{JSE}})^\top + \Delta^{\text{JSE}}, \tag{35}$$

where $\eta^2, h^{\text{JSE}}, \Delta^{\text{JSE}}$ are estimators of ζ^2, b, Δ described further in Section 4. The JSE estimators are designed for large $p >> n$ and correct for high-dimensional statistical bias in the sample eigenvectors. Suitable parameters are $p = 1000$, $n = 126$ as in the empirical illustrations in the next section. This is a single-factor model where the normalized factor loadings h^{JSE} are asymptotically good estimates of the unknown population factor exposure unit vector $b = \beta/|\beta|$ responsible for generating the observed returns via (33).

The remaining two estimators $\Sigma^{\text{MJSE}}, \Sigma^{\text{MS}}$ are single-index market models in the manner of Sharpe [1963] and as used by Clarke et al. [2011].

Given any data-driven estimator $\hat{\Sigma}$ of Σ, and cap-weighted market portfolio w_M, we may define the estimated market variance and market beta by

$$\sigma_M^2 = w_M^\top \hat{\Sigma} w_M \tag{36}$$

and

$$\beta^M = \frac{\hat{\Sigma} w_M}{\sigma_M^2}. \tag{37}$$

From these, we form a single index market covariance matrix $\Sigma^{M,\hat{\Sigma}}$ depending on w_M and $\hat{\Sigma}$:

$$\Sigma^{M,\hat{\Sigma}} = \sigma_M^2 \beta^M (\beta^M)^\top + \Delta^M, \tag{38}$$

where

$$\Delta^M = diag((\delta_i^M)^2), \quad (\delta_i^M)^2 = S_{ii} - (\beta_i^M)^2 \sigma_M^2, \tag{39}$$

or, equivalently,

$$\Sigma^{M,\hat{\Sigma}} = \zeta_M^2 b^M (b^M)^\top + \Delta^M \tag{40}$$

where

$$\zeta_M^2 = \sigma_M^2 |\beta^M|^2, \quad b^M = \beta^M / |\beta^M|. \tag{41}$$

We examine the choices $\hat{\Sigma} = \Sigma^{\text{JSE}}$ and $\hat{\Sigma} = S$, and define

$$\Sigma^{\text{MJSE}} = \Sigma^{M,\Sigma^{\text{JSE}}} \tag{42}$$

$$\Sigma^{\text{MS}} = \Sigma^{M,S}. \tag{43}$$

It can be shown that Σ^{MJSE} and Σ^{JSE} as asymptotically consistent (as $p \to \infty$) in a sense described in Section 4.2 below.

2.2 Empirical example for a single index model In this section we apply our results in an empirical illustration using daily returns of the top $p = 1000$ US stocks by market capitalization for the period January 3, 2022 to July 1, 2022.[4]

From the market capitalizations $mc(i), i = 1, \ldots, p$, at the beginning of the period, we determine the market portfolio w_M by

$$w_M(i) = \frac{mc(i)}{\sum_{i=1}^p mc(i)}. \tag{44}$$

We then use the formulas of section 2.1 to determine three choices[5] of the data-driven covariance estimator for computing the LOMV portfolio: Σ^{JSE}, Σ^{MJSE}, and Σ^{MS}.

The outcomes for the resulting long only portfolio size, market factor variance, and estimated betas are summarized in Table 1. For $\Sigma^{\text{JSE}} = \eta^2 h^{\text{JSE}} h^{\text{JSE}\top} + \Delta^{\text{JSE}}$, the market quantities σ_M^2 and β^M are undefined and h^{JSE} is scaled as a unit vector by convention.

estimator	k	$\sigma_M^2(\%)$	mean(β^M)
Σ^{MJSE}	65	5.8	1.08
Σ^{MS}	51	6.2	1.13
Σ^{JSE}	66	*	*

Table 1: Outcomes for each of three choices of covariance estimator. The value k is the number of active assets in the long-only minimum variance portfolio (out of 1000). σ_M^2 and β^M are the derived parameters according to equations (36) and (37) but does not apply to the last row.

[4]Returns were gathered from the WRDS database, then cleaned and centered before use. We removed one smaller stock of a firm with artificially low volatility due to an imminent merger, and added the next largest asset to keep the total at 1,000.

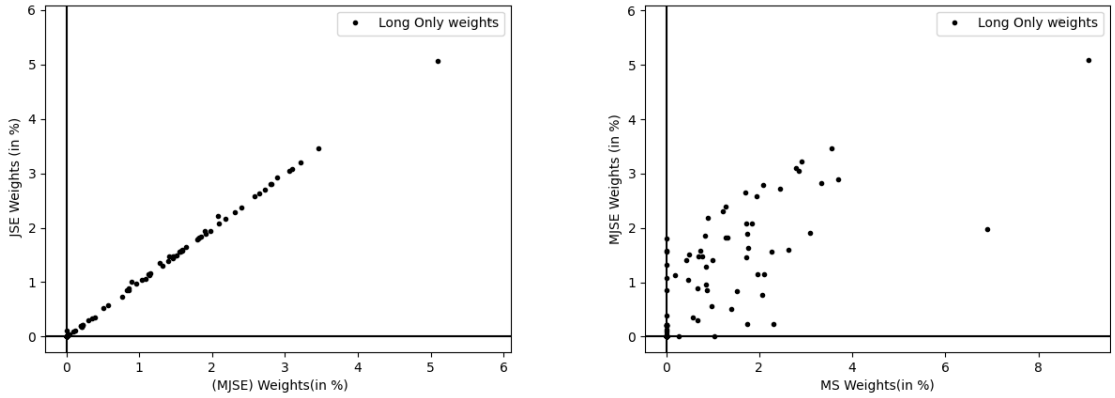
[5]Clarke et al. [2011] use a different choice of single-index market covariance matrix by taking $\hat{\Sigma}$ to be a Ledoit-Wolf shrinkage estimator, see Ledoit and Wolf [2004]. The empirical outcomes are similar to the ones reported here.

Table 2 shows that the three methods mostly agree on which active assets should be included in the long-only minimum variance portfolio.

portfolio	MJSE	MS	JSE	MJSE ∩ MS	MJSE ∩ JSE	JSE ∩ MS
# assets	65	51	66	49	65	49

Table 2: The number of assets that each of the three estimated minimum variance portfolios have in common, out of 1000 total assets considered.

Figures 1a and 1b show scatterplots demonstrating that the portfolios for the market model Σ^{MJSE} and the statistical factor model Σ^{JSE} are almost the same in this experiment, but the Σ^{MS} portfolio noticeably differs.



(a) Weights of MJSE vs JSE portfolios.

(b) Weights of MS vs MJSE portfolios.

Figure 1: Comparison of portfolio asset weights for the three portfolios MJSE, JSE, and MS, plotted in percent.

Figures 2, 3, and 4 below illustrate the relationships between estimated market beta, portfolio weight, and idiosyncratic risk for the MSJE portfolio. (Plots look similar for the other portfolios.)

The betas range between -0.05 and 3.62 with the only negative beta being approximately -0.059. The maximum beta in the long-only portfolio is 0.281. The idiosyncratic risk ranges from 0 to 46%. Figures 2 and 3 show the 1000 individual asset weights under the single-factor model for both the long-short (blue dots) and the long-only (black dots) minimum-variance portfolio plotted against market beta and delta. The weights of the active assets in the long-only portfolio had a maximum value of 5.81%.

Of interest is the fact that only 65 of the 1000 securities were active in the long-only portfolio. In addition to having lower beta, assets with low idiosyncratic risk are more likely to be active in the long-only portfolio. We also see that assets with lower idiosyncratic risk tend to have higher absolute value weights in the long-short portfolio.

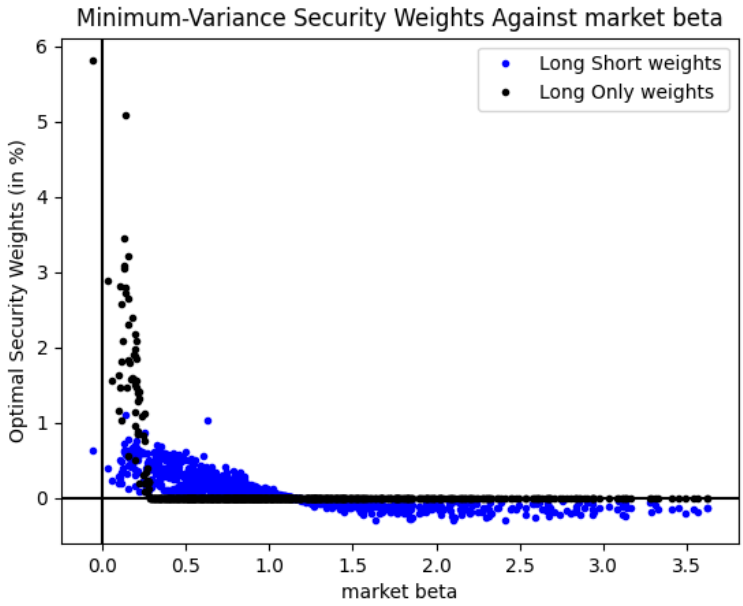


Figure 2: MJSE portfolio weights against market beta for the top 1000 US stocks, estimated for daily returns in the first half of 2022.

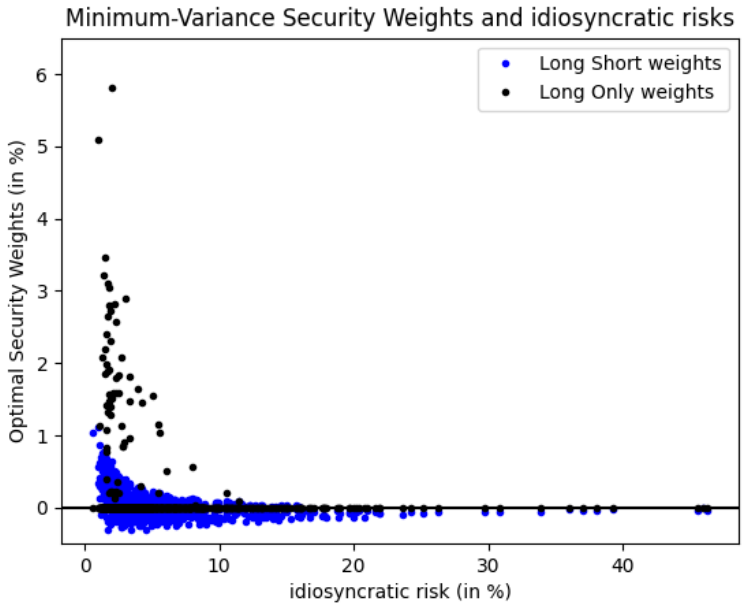


Figure 3: MJSE portfolio weights against specific risk for the top 1000 US stocks, from daily returns in the first half of 2022.

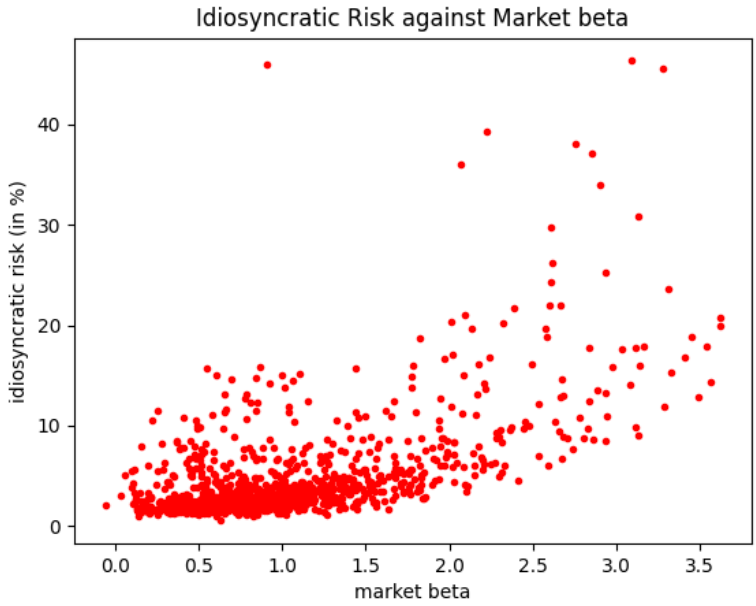


Figure 4: MJSE market beta against specific risk for the top 1000 US stocks in the first half of 2022.

2.3 Empirical example for a two-factor model To illustrate the hyperplane separation of factor loadings described in Theorem 3, we fit a two-factor model ($q = 2$) to the same daily returns of the 1,000 stocks used in Section 2.2. Starting from the leading two eigenvalue-eigenvector pairs of the sample covariance matrix, as described further in Section 5, we apply the James-Stein-Markowitz (JSM) multifactor shrinkage method of Shkolnik et al. [2024] to obtain the covariance estimate

$$\Sigma^{\text{JSM}} = B\Omega B^{\top} + \Delta, \tag{45}$$

where B is a $2 \times p$ matrix with orthogonal columns, obtained via shrinkage from the leading two sample eigenvectors; Δ is a diagonal matrix of specific variances; and Ω is a 2×2 diagonal matrix. The row B_i of asset i is that asset's exposure vector to the two factors. The resulting long-only minimum variance portfolio is computed by means of the open source convex optimization package cvxpy (www.cvxpy.org).

The results are displayed in the left hand plot of Figure 5 below, in which the horizontal axis shows the exposure to the leading (market) factor (the first component of B_i), and the vertical axis to the second factor. Orange indicates assets included in the LOMV portfolio based on the model, and blue indicates excluded assets. The solid red line is the hyperplane of Theorem 3 separating the active and inactive assets in the LOMV portfolio, while the dotted red line indicates the single-index model threshold value that would apply for the same data with a one-factor model. We observe a relatively small number of assets in the LOMV portfolio.

The right-hand plot in Figure 5 shows a close-up of the LOMV portfolio's active assets with the portfolio weights coded as marker size and color. The largest weight is 10%. The

portfolio weight of an asset is proportional[6] to the ratio d_i/δ_i^2, where δ_i^2 is the specific variance of asset i, and d_i is the perpendicular distance from the asset's beta point B_i to the separating hyperplane. Although the active asset portfolio weights shown in Figure 5 generally vary inversely with the distance from the separating hyperplane, unusually large or small specific variances can play a dominant role, as is the case for three assets with large exposure to the second factor.

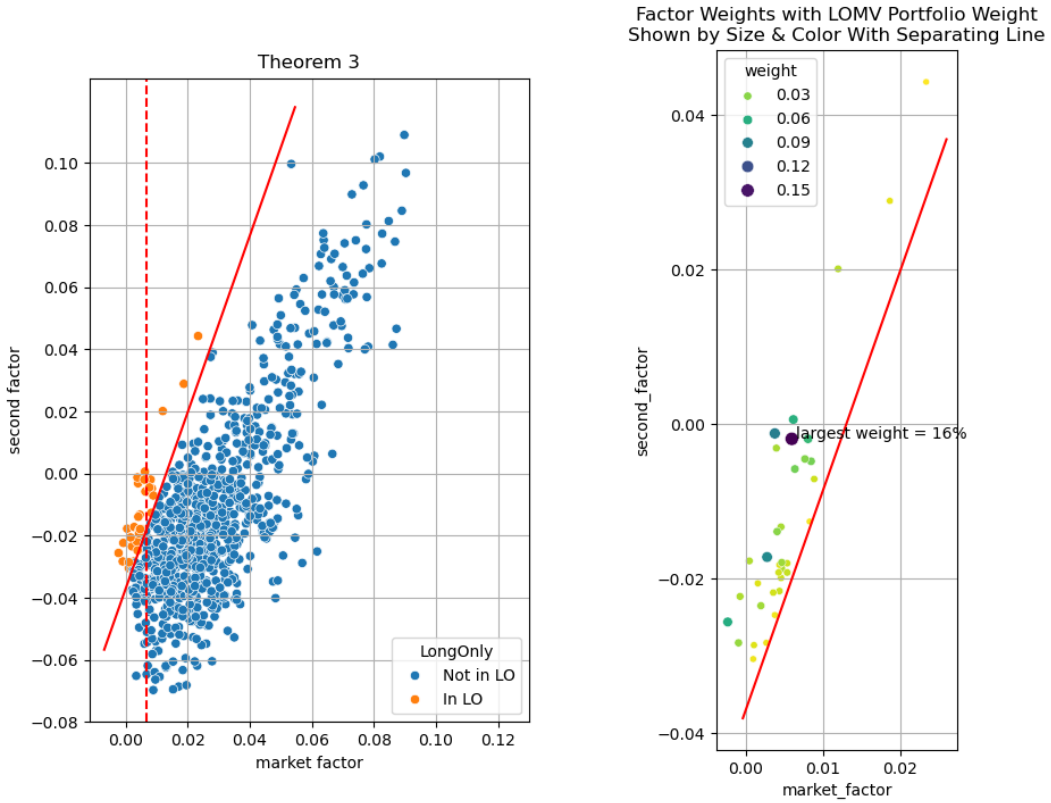


Figure 5: 2-vector exposures for each asset from a statistical 2-factor JSM return covariance matrix, plotted for each of the 1000 assets. On the right is a close-up showing portfolio weights.

An interesting comparison with the portfolio resulting from a single-index model is evident by examining the vertical dotted red line in the left-hand plot, which, as stated above, is the one-factor threshold for the same returns data. The 1-factor portfolio excludes a few of the assets with higher exposure to both factors and includes significantly more with negative exposure to the second factor. For the two-factor model portfolio, exposure to the second factor can compensate for a fairly significant positive exposure to the first factor, such as for the two uppermost orange assets in left-hand plot. These assets are significantly into the

[6]With a proportionality constant that is the same for all assets. This follows from a computation using the Woodbury identity.

excluded region defined by the dotted line. Overall, the 2-factor model includes 41 assets, 24 fewer than the 65 assets of one-factor model.

The angle between the two lines could be considered a measure of the importance of the second factor in determining active long-only assets. The angle here is 21.5% of a right angle, which is relatively significant.

3 Proofs

3.1 Proof of Theorem 1 Notation: $\mathbf{1}_p$ is the p-dimensional vectors of all ones, and similarly $\mathbf{0}_p$ for all zeros.

The solution $w = w^L$ of our constrained optimization problem (2) satisfies the well-known Karush-Kuhn-Tucker (KKT) conditions[7], which are a set of equations in w, an auxiliary p-vector λ, and an auxiliary scalar ν (the Lagrange multipliers) as follows:

$$2\Sigma w - \lambda + \nu \mathbf{1}_p = \mathbf{0}_p \tag{46}$$
$$w^\top \mathbf{1}_p = 1 \tag{47}$$
$$\lambda_i w_i = 0 \ i = 1, 2, ..., p \tag{48}$$
$$\lambda_i \geq 0, w_i \geq 0 \ i = 1, 2, ..., p. \tag{49}$$

These KKT conditions include $2p + 1$ equations (46) – (48) in the $2p + 1$ unknowns

$$w_1, \ldots, w_p, \lambda_1, \ldots, \lambda_p, \nu,$$

along with $2p$ inequality constraints (49).

Define the (necessarily non-empty) set

$$K = \{i \leq p : w_i > 0\}. \tag{50}$$

Let $k > 0$ denote the cardinality of K. For any vector $x \in \mathbf{R}^p$, denote by $x^K \in \mathbf{R}^k$ the k-dimensional vector obtained from x by deleting the entries x_j for all $j \notin K$. Likewise, for any $p \times p$ matrix M, denote by M^K the $k \times k$ principal submatrix obtained from M by deleting all the rows and columns with indices outside K.

Since Σ is symmetric positive definite, so is the submatrix Σ^K, and hence Σ^K is invertible. Further,

$$w^\top \Sigma w = (w^K)^\top \Sigma^K w^K. \tag{51}$$

Since $\lambda_i = 0$ for all $i \in K$, taking the k rows of equation (46) corresponding to indices in K tells us

$$2\Sigma^K w^K + \nu \mathbf{1}_k = \mathbf{0}_k, \tag{52}$$

where here the vectors $\mathbf{1}_k$ and $\mathbf{0}_k$ are k-dimensional. Multiplying (52) on the left by $\mathbf{1}_k^\top (\Sigma^K)^{-1}$ and using $\mathbf{1}_k^\top w^K = 1$, we obtain

$$\nu = \frac{-2}{\mathbf{1}_k^\top (\Sigma^K)^{-1} \mathbf{1}_k} \tag{53}$$

[7]See, for example, Beck [2023].

and therefore

$$w^K = \frac{(\Sigma^K)^{-1}\mathbf{1}_k}{\mathbf{1}_k^\top(\Sigma^K)^{-1}\mathbf{1}_k}, \tag{54}$$

or, equivalently,

$$w^L = \frac{(\Sigma^{K,0})^+\mathbf{1}_p}{\mathbf{1}_p^\top(\Sigma^{K,0})^+\mathbf{1}_p}. \tag{55}$$

3.2 Proof of Theorem 2 Part 1 of the theorem follows solely from the monotonicity of the sequence $\{\beta_i\}$. For convenience, define

$$C_i = \sum_{j=1}^{i} \frac{\beta_j}{\delta_j^2}. \tag{56}$$

It is easy to verify, for all $i = 1, \ldots, p-1$, that

$$R_{i+1} - R_i = (-\beta_{i+1} + \beta_i)C_i. \tag{57}$$

Since $(-\beta_{i+1} + \beta_i)$ is always non-positive, $R_{i+1} - R_i \leq 0$ when $C_i \geq 0$ and $R_{i+1} - R_i \geq 0$ when $C_i \leq 0$.

Let $s = \min\{i : C_i > 0\}$, $1 \leq s \leq p$. If $j < s$, then $C_j \leq 0$ by definition of s, so $R_{j+1} - R_j \geq 0$. Since $\beta_s > 0$, C_i is increasing for $i \geq s$, so if $j \geq s$ then $C_j > 0$, and we have $R_{j+1} - R_j \leq 0$. This establishes the conclusion of part 1.

We proceed to Part 2. Our goal now is to determine $K = \{i \leq p : w_i^L > 0\}$ in explicit form in term of the parameters σ, β, δ of the problem.

Let $k = |K|$ and define

$$\ell = \max\{i \leq p : R_i > 0\}. \tag{58}$$

We will complete the proof by establishing

$$k = \ell \text{ and } K = \{1, 2, \ldots, k\}.$$

Let $\beta^K \in \mathbf{R}^k$ and the $k \times k$ matrix Σ^K be determined from K as before, obtained from β and Σ by deleting rows or rows and columns corresponding to indices outside K.

By the Woodbury identity,

$$(\Sigma^K)^{-1} = diag(\frac{1}{(\delta^K)^2}) - \frac{(\frac{\beta^K}{(\delta^K)^2})(\frac{\beta^K}{(\delta^K)^2})^\top}{\frac{1}{\sigma^2} + (\frac{\beta^K}{(\delta^K)^2})^\top\beta^K}$$

where $\frac{1}{(\delta^K)^2} = [\frac{1}{\delta_j^2} : j \in K]^\top$ and $\frac{\beta^K}{(\delta^K)^2} = [\frac{\beta_j}{\delta_j^2} : j \in K]^\top$.

This means

$$(\Sigma^K)^{-1}\mathbf{1}_k = \frac{1}{(\delta^K)^2} - \frac{(\frac{\beta^K}{(\delta^K)^2})(\frac{\beta^K}{(\delta^K)^2})^\top\mathbf{1}_k}{\frac{1}{\sigma^2} + (\frac{\beta^K}{(\delta^K)^2})^\top\beta^K}. \tag{59}$$

Now if $i \in K$, then

$$((\Sigma^K)^{-1}\mathbf{1}_k)_i = \frac{1}{\delta_i^2} - \frac{\frac{\beta_i}{\delta_i^2}\sum_{j\in K}\frac{\beta_j}{\delta_j^2}}{\frac{1}{\sigma^2} + \sum_{j\in K}\frac{\beta_j^2}{\delta_j^2}} > 0. \tag{60}$$

Clearing the positive denominators, we obtain

$$\frac{1}{\sigma^2} + \sum_{j \in K} \frac{\beta_j^2}{\delta_j^2} - \beta_i \sum_{j \in K} \frac{\beta_j}{\delta_j^2} > 0 \tag{61}$$

or

$$B_K > \beta_i C_K \tag{62}$$

where

$$B_K = \frac{1}{\sigma^2} + \sum_{j \in K} \frac{\beta_j^2}{\delta_j^2}, \quad C_K = \sum_{j \in K} \frac{\beta_j}{\delta_j^2}. \tag{63}$$

Now suppose instead $i \notin K$, so $w_i = 0$. Since $\Sigma w = \sigma^2 \beta \beta^\top w + diag(\delta^2) w$ and $w_j = 0$ for all $j \notin K$, we have

$$(\Sigma w)_i = \sigma^2 \beta_i \sum_{j \in K} \beta_j w_j. \tag{64}$$

Conditions (46) and (49) tell us

$$0 \leq \lambda_i = 2(\Sigma w)_i + \nu, \tag{65}$$

or, using (64),

$$0 \leq 2\sigma^2 \beta_i \sum_{j \in K} \beta_j w_j + \nu. \tag{66}$$

For $j \in K$, we have, from (54),

$$w_j = \frac{((\Sigma^K)^{-1} \mathbf{1}_k)_j}{\mathbf{1}_k (\Sigma^K)^{-1} \mathbf{1}_k}. \tag{67}$$

Using this, substituting for ν with (53), and multiplying through by $\mathbf{1}_k (\Sigma^K)^{-1} \mathbf{1}_k / 2$ gives us

$$0 \leq \sigma^2 \beta_i \sum_{j \in K} \beta_j ((\Sigma^K)^{-1} \mathbf{1}_k)_j - 1. \tag{68}$$

By (60),

$$((\Sigma^K)^{-1} \mathbf{1}_k)_j = \frac{1}{\delta_j^2} - \frac{\beta_j}{\delta_j^2} \frac{C_K}{B_K}. \tag{69}$$

Using this in the previous inequality, we have

$$\begin{aligned} 0 &\leq \sigma^2 \beta_i \sum_{j \in K} \frac{\beta_j}{\delta_j^2} (1 - \frac{\beta_j C_K}{B_K}) - 1 && (70) \\ &= \sigma^2 \beta_i \big(C_K - \frac{C_K}{B_K} (B_K - \frac{1}{\sigma^2}) \big) - 1 && (71) \\ &= \sigma^2 \beta_i \frac{C_K}{\sigma^2 B_K} - 1 = \beta_i \frac{C_K}{B_K} - 1, && (72) \end{aligned}$$

or

$$B_K \leq \beta_i C_K. \tag{73}$$

Combining (62) and (73), we have established the following

Lemma 1. *With B_K and C_K as defined above, $i \in K$ if and only if $B_K > \beta_i C_K$.*

Corollary 3. $C_K \neq 0$.

Proof of Corollary. If $k = p$, then $C_K \neq 0$ is our standing assumption. Otherwise there exists $j \notin K$, so by Lemma 1 $B_K \leq \beta_j C_K$. But $B_K > 0$, so again C_K cannot be zero. □

Lemma 2. *Recall k is the cardinality of K. If $C_K > 0$, then $K = \{1, 2, \ldots, k\}$. If $C_K < 0$, then $K = \{p - k + 1, \ldots, p\}$.*

Proof of lemma. Recall that the β_i are arranged in increasing order. For all $i \in K$ and $j \notin K$, we have

$$\beta_i C_K < B_K \leq \beta_j C_K, \tag{74}$$

hence

$$\beta_i C_K < \beta_j C_K. \tag{75}$$

If $C_K > 0$, this means that $\beta_i < \beta_j$ for all $i \in K, j \notin K$, and hence $i < j$ for all $i \in K, j \notin K$. Therefore K must be an initial segment of the sequence $\{1, 2, \ldots, p\}$.

Similarly, if $C_K < 0$, then the reverse inequality is true, and K must be a terminal segment of $\{1, 2, \ldots, p\}$. □

Next, define

$$C_P = \sum_{j=1}^{p} \frac{\beta_j}{\delta_j^2}, \tag{76}$$

and recall $C_P > 0$ by our standing assumption.

Lemma 3. $C_K > 0$.

Proof. If $k = p$ then $C_K = C_P$ and there is nothing to prove, so consider the case $k < p$. We know that $C_K \neq 0$. Suppose for contradiction that $C_K < 0$. By Lemma 2, this means that $K = \{p - k + 1, \ldots, p\}$.

Note

$$C_P = \sum_{j=1}^{p-k} \frac{\beta_j}{\delta_j^2} + C_K. \tag{77}$$

Now $0 > C_K = \sum_{j=p-k+1}^{p} \frac{\beta_j}{\delta_j^2}$, so at least one of the terms of this sum must be negative. Then, since the beta sequence is increasing, $\beta_{p-k+1} < 0$ and $\beta_j < 0$ for all $j \leq p - k$, and hence

$$\sum_{j=1}^{p-k} \frac{\beta_j}{\delta_j^2} < 0. \tag{78}$$

This forces $C_P < 0$ via (77), a contradiction. Hence we must have $C_K > 0$ and $C_K = \sum_{j=1}^{k} \frac{\beta_j}{\delta_j^2}$. □

Lemma 4. *Under our standing assumption, for the long-only solution w of problem (2),*

$$w_i > 0 \text{ if and only if } \beta_i < \frac{B_K}{C_K} = \frac{\frac{1}{\sigma^2} + \sum_{j=1}^{k} \frac{\beta_j^2}{\delta_j^2}}{\sum_{j=1}^{k} \frac{\beta_j}{\delta_j^2}}. \tag{79}$$

Proof. From Lemma 3, $C_K > 0$. By Lemma 2, $K = \{1, 2, \ldots, k\}$. The result follows from Lemma 1. □

By Lemmas 2 and 3, we have established that

$$K = \{1, 2, \ldots, k\}. \tag{80}$$

It remains to show that $k = \ell$. Recall $R_1 = 1/\sigma^2$ and, for $i = 2, \ldots, p$,

$$R_i = \frac{1}{\sigma^2} + \sum_{j=1}^{i-1} \frac{\beta_j}{\delta_j^2}(\beta_j - \beta_i). \tag{81}$$

Now, by Lemma 4, $\beta_k < B_K/C_K \leq \beta_{k+1}$. Also,

$$\begin{aligned} R_k &= \frac{1}{\sigma^2} + \sum_{j=1}^{k} \frac{\beta_j^2}{\delta_j^2} - \beta_k \sum_{j=1}^{k} \frac{\beta_j}{\delta_j^2} && (82) \\ &= B_K - \beta_k C_K > 0. && (83) \end{aligned}$$

Likewise

$$R_{k+1} = B_K - \beta_{k+1} C_K \leq 0. \tag{84}$$

By part 1 of the Theorem, the sequence $\{R_i\}$ crosses zero at most once, so this establishes

$$k = \max\{i \leq p : R_i > 0\} = \ell, \tag{85}$$

completing the proof of Part 2. □

3.3 Proof of Theorem 3 Let $w = w^L$ be the solution of the long-only problem for

$$\Sigma = B\Omega B^\top + \Delta, \tag{86}$$

and $K = \{i \leq p : w_i^L > 0\}$, with $\ell = |K|$. For any p-vector v, for purposes of this proof we adopt the notation that v^K as defined before, and v' is the complementary $(p-\ell)$-vector obtained from v by deleting all the coordinates in K; B' is obtained by deleting all the rows labeled by entries in K, and B^K by deleting the rows not in K.

First recall that $i \notin K$ implies $w_i = 0$, and hence

$$(\Sigma w)' = \big((B\Omega B^\top + \Delta)w\big)' = (B\Omega B^\top w)' = B'\Omega (B^K)^\top w^K. \tag{87}$$

Recall equations (46) and (49):

$$2\Sigma w - \lambda + \nu \mathbf{1}_p = \mathbf{0}_p; \quad \lambda_i \geq 0, w_i \geq 0,\ i = 1, \ldots, p.$$

Therefore

$$\mathbf{0}_{p-\ell} \leq \lambda' = 2(\Sigma w)' + \nu \mathbf{1}_{p-\ell} \tag{88}$$

and hence

$$\mathbf{0}_{p-\ell} \leq 2B'\Omega(B^K)^\top w^K + \nu \mathbf{1}_{p-\ell} \tag{89}$$

using equation (87).

Substituting using (53) and (54), clearing denominators, and dividing by 2, we obtain

$$\mathbf{0}_{p-\ell} \leq B'\Omega(B^K)^\top(\Sigma^K)^{-1}\mathbf{1}_\ell - \mathbf{1}_{p-\ell} = B'h^K - \mathbf{1}_{p-\ell} \tag{90}$$

with

$$h^K = \Omega(B^K)^\top(\Sigma^K)^{-1}\mathbf{1}_\ell$$

as defined in (26) of Theorem 3.

Thus we have established that for every $i \notin K$,

$$1 \leq B_i h^K. \tag{91}$$

Conversely, recall

$$\Sigma^K = B^K\Omega(B^K)^\top + \Delta^K. \tag{92}$$

We therefore have

$$B^K h^K = B^K\Omega(B^K)^\top(\Sigma^K)^{-1}\mathbf{1}_\ell = (\Sigma^K - \Delta^K)(\Sigma^K)^{-1}\mathbf{1}_\ell = \mathbf{1}_\ell - \Delta^K(\Sigma^K)^{-1}\mathbf{1}_\ell. \tag{93}$$

From equation (54), $(\Sigma^K)^{-1}\mathbf{1}_\ell = (\mathbf{1}_\ell^\top(\Sigma^K)^{-1}\mathbf{1}_\ell)w^K$, and hence

$$B^K h^K = \mathbf{1}_\ell - (\mathbf{1}_\ell^\top(\Sigma^K)^{-1}\mathbf{1}_\ell)\Delta^K w^K < \mathbf{1}_\ell, \tag{94}$$

since the entries of $(\mathbf{1}_\ell^\top(\Sigma^K)^{-1}\mathbf{1}_\ell)\Delta^K w^K$ are all positive. We conclude that

$$B_i h^K < 1 \tag{95}$$

for all $i \in K$. □

4 Single factor beta estimation

4.1 The JSE estimator Covariance matrix shrinkage estimators, e.g. Ledoit and Wolf [2004], have enjoyed wide adoption for various problems when estimating large-dimensional covariance matrices from data, and typically take the form

$$\Sigma = \alpha S + (1-\alpha)T, \tag{96}$$

where S is a sample covariance matrix, T is a suitable target matrix, such as a scalar matrix, and $\alpha \in (0,1)$ is defined to minimize some error function.

However, in cases where we are working within the structure of a factor model, we are primarily interested in estimating the leading covariance eigenvector(s), from which an estimated factor model can be defined. Eigenvector shrinkage is the subject of a recent stream of research in Goldberg et al. [2020], Goldberg and Kercheval [2023], Goldberg et al. [2022, 2025], Shkolnik [2022]. We call it JSE (James-Stein for Eigenvectors) because the shrinkage

formulas themselves turn out to be close analogs of the classical James-Stein shrinkage formulas for estimation of multivariate means. The theory provides asymptotic improvements in the HL asymptotic regime corresponding to the limit as the dimension $p \to \infty$ with the number of samples n fixed.

In this section we describe how to compute the JSE estimator of the leading eigenvector.

As in Section 2.1, we consider a single factor model of returns,

$$r = \beta f + \epsilon. \tag{97}$$

The covariance matrix of r is given by (10):

$$\Sigma = \sigma^2 \beta\beta^T + \Delta.$$

Setting $\zeta^2 = ||\beta||^2\sigma^2$ and $b = \beta/||\beta||$, the single index model may be written

$$\Sigma = \zeta^2 bb^\top + \Delta, \tag{98}$$

where recall $\Delta = diag(\delta_1^2, \ldots, \delta_p^2)$. Estimating Σ requires estimating the scalar ζ^2, the unit vector b, and the vector $(\delta_1^2, \ldots, \delta_p^2)$ of idiosyncratic variances.

Suppose we observe a time series of n returns of each of p assets, and $p >> n$ as might typically be the case when n is limited by non-stationarity or data availability. The observations can be summarized by a $p \times n$ data matrix Y, and the resulting sample covariance matrix is

$$S = YY^\top/n. \tag{99}$$

Let λ^2 denote the leading eigenvalue of S, and let

$$\ell^2 = \frac{tr(S) - \lambda^2}{n-1} \tag{100}$$

be the average of the non-zero eigenvalues that are less than λ^2, where $tr(S)$ denotes trace. We can think of

$$\eta^2 = \lambda^2 - \ell^2$$

as the average leading sample eigengap, and it turns out that η^2 is an unbiased approximation of ζ^2.

To approximate the vector $b = \beta/||\beta||$, we could select the leading sample eigenvector h of S. However, for fixed n, the asymptotic limit in p of the cosine of the angle beetween h and b is positive. In fact it is equal to

$$\big(1 + \frac{\delta^2}{nB^2\sigma^2}\big)^{-1} < 1,$$

where B^2 is the limit of $||\beta||^2/p$ and δ^2 is the limit of $(1/p)\sum_{i=1}^p \delta_i^2$ as $p \to \infty$.

A definite improvement is obtained with the James-Stein eigenvector shrinkage estimator, h^{JSE}, defined as follows. Let h_1 denote the projection of h onto the line spanned by $\mathbf{1}$. Define the shrinkage constant

$$c = \frac{\ell^2}{\lambda^2(1 - ||h_1||^2)} \tag{101}$$

and

$$H = ch_1 + (1-c)h. \tag{102}$$

Then

$$h^{\mathrm{JSE}} = H/||H||. \tag{103}$$

The unit vector h^{JSE} is obtained from h by correcting a concentration of measure effect that pushes h farther away from $\mathbf{1}$ than b.

Recalling that the ith diagonal element s_{ii} of S is the sample variance of the ith asset, we estimate the ith idiosyncratic variance as

$$\hat{\delta}_i^2 = s_{ii} - \eta^2 (h_i^{\mathrm{JSE}})^2. \tag{104}$$

Our estimated non-singular covariance matrix Σ^{JSE} is now

$$\Sigma^{\mathrm{JSE}} = \eta^2 h^{\mathrm{JSE}} (h^{\mathrm{JSE}})^\top + diag(\hat{\delta}_1^2, \ldots, \hat{\delta}_p^2). \tag{105}$$

4.2 Consistency of single index estimators When estimating betas from market data as in Section 2.1, there are three covariance matrices in the picture:

1. the unobserved population covariance
$$\Sigma = \sigma^2 \beta\beta^\top + \Delta,$$
2. the covariance estimated from observed returns
$$\Sigma^{\mathrm{JSE}} = \eta^2 h^{\mathrm{JSE}} (h^{\mathrm{JSE}})^\top + \Delta^{\mathrm{JSE}},$$
3. the market covariance incorporating the observed market portfolio w_M
$$\Sigma^M = \sigma_M^2 \beta^M (\beta^M)^\top + \Delta^M$$
with notation defined in (36), (37), and (39).

There is a consistency question with this approach, because $w_M^\top \Sigma^M w_M \neq \sigma_M^2$. Further, the beta factors for the two estimators disagree: $(\eta/\sigma_M) h^{\mathrm{JSE}} \neq \beta^M$.

However, a single factor estimator like Σ^{JSE} has the benefit that the inconsistency above vanishes for a large number of assets, as $p \to \infty$.

Direct computation establishes the following

Theorem 4. *Consider a sequence in p of p-dimensional population covariance models*

$$\Sigma = \sigma^2 \beta\beta^\top + \Delta$$

where σ^2 is fixed, Δ is bounded in p, and $|\beta|^2/p$ tends to a positive finite limit. Suppose the number of observations used to determine the sample covariance matrix is fixed independent of p.

Then, with Σ^M computed using $\hat{\Sigma} = \Sigma^{\mathrm{JSE}}$, we have, as $p \to \infty$,

$$|b^M - h^{\mathrm{JSE}}| \to 0 \tag{106}$$
$$|\sigma_M^2 |\beta^M|^2 - \eta^2|/p \to 0 \tag{107}$$
$$|\delta_i^M - \Delta_i^{\mathrm{JSE}}| \to 0. \tag{108}$$

This means that the market beta, if computed with Σ^{JSE}, is asymptotically the same as the leading factor in the single index covariance matrix that defines it.

5 Multifactor JSM estimation

The multifactor JSM estimator is obtained from a PCA estimate by an appropriate shrinkage of the q-dimensional factor subspace toward a suitable shrinkage target. It is suitable when $p >> n$; complete discussion of this method appears in Shkolnik et al. [2024].

In this section we summarize the method, which is the same for any q. The returns model is

$$r = Bf + \epsilon, \tag{109}$$

where f is a mean-zero random q-vector of returns to q risk factors, ϵ is a mean zero random p-vector of specific returns whose components are independent of each other and of f, and B is an unknown $p \times q$ parameter matrix of sensitivities of the securities to the factors.

With a time series of n i.i.d. returns, $q < n < p$, let R denote the $p \times n$ matrix of centered returns data. For the sample covariance matrix $S = RR^\top/n$ with rank $n_+ \leq n$ and positive eigenvalues $\lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2 \geq \cdots \geq \lambda_{n_+}^2$, take the spectral decomposition

$$S = \sum_{i=1}^{n_+} \lambda_i^2 h_i h_i^\top = HH^\top + N, \tag{110}$$

where h_i is the unit eigenvector corresponding to λ_i^2, H is the $p \times q$ matrix with columns $\lambda_1 h_1, \ldots, \lambda_q h_q$, and $N = S - HH^\top$. Then form the specific variance estimate $\Delta = diag(N)$ to be the diagonal matrix with the same diagonal as N.

The q-factor PCA covariance estimator is $\Sigma^{\text{PCA}} = HH^\top + \Delta$, but this can be improved by shrinking H to obtain

$$\Sigma^{\text{JSM}} = H_{\text{JSM}} H_{\text{JSM}}^\top + \Delta \tag{111}$$

as follows.

From the re-weighted data matrix $R_w = \Delta^{-1/2} R$, we can recompute the $p \times q$ leading eigenvector matrix H via

$$R_w R_w^\top/n = H_w H_w^\top + N_w \tag{112}$$

and then let $\bar{H} = \Delta^{1/2} H_w$. Define a $p \times q$ shrinkage target

$$M = \mathbf{1}_p (\mathbf{1}_p^\top \Delta^{-1} \mathbf{1}_p)^{-1} \mathbf{1}_p^\top \Delta^{-1} \bar{H}. \tag{113}$$

Then H_{JSM} is defined by linear shrinkage toward M:

$$H_{\text{JSM}} = \bar{H} C + M(I - C) \tag{114}$$

where

$$C = I - \nu^2 J^{-1}, \quad J = (\bar{H} - M)^\top \Delta^{-1} (\bar{H} - M), \tag{115}$$

and

$$\nu^2 = \frac{trace(\Delta)}{n_+ - q}. \tag{116}$$

The format

$$H_{\text{JSM}} = B \Omega B^\top + \Delta \tag{117}$$

is obtained by setting the columns of B to be the ordered unit eigenvectors of $H_{\text{JSM}} H_{\text{JSM}}{}^\top$, and Ω the diagonal matrix of corresponding eigenvalues.

References

Amir Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with Python and MATLAB*. SIAM, 2nd edition, 2023.

R. Clarke, H. De Silva, and S. Thorley. Minimum-variance portfolio composition. *The Journal of Portfolio Management*, Winter:31–45, 2011.

L. R. Goldberg, H. Gurdogan, and A. Kercheval. Portfolio optimization via strategy-specific eigenvector shrinkage. *Finance and Stochastics*, https://doi.org/10.1007/s00780-025-00566-4, 2025.

Lisa R. Goldberg and Alec N. Kercheval. James-Stein for the leading eigenvector. *Proceedings of the National Academy of Sciences*, 120(2), 2023.

Lisa R. Goldberg, Alex Papacinicolau, Alex Shkolnik, and Simge Ulucam. Better betas. *The Journal of Portfolio Management*, 47(1):119–136, 2020.

Lisa R. Goldberg, Alex Papanicolau, and Alex Shkolnik. The dispersion bias. *SIAM Journal of Financial Mathematics*, 13(2):521–550, 2022.

O. Ledoit and M. Wolf. Honey, I shrunk the covariance matrix. *The Journal of Portfolio Management*, Fall:110–119, 2004.

William F. Sharpe. A simplified model for portfolio analysis. *Management Science*, 9(2): 277–293, Jan. 1963.

A. Shkolnik, A. Kercheval, H. Gurdogan, L. Goldberg, and H. Bar. Portfolio selection revisited. *Annals of Operations Research*, 2024.

Alex Shkolnik. James-Stein estimation of the first principal component. *Stat*, 11(1), 2022.