Explicit Solution for Position Constrained Minimum Variance Portfolios

Alex Bernstein
abernstein@ucsb.edu
joint work with Alexander Shkolnik

Seminar at University of California, Berkeley
March 3, 2020

Department of Statistics & Applied Probability
University of California, Santa Barbara

Overview of this Talk

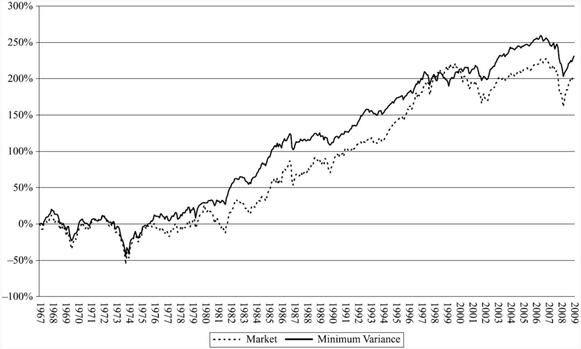
History of Quadratic Programming Let Σ be a covariance matrix with a factor structure. Our main result is a semi-explicit solution to the program

$$
\begin{aligned}
&\min_{x \in \mathbb{R}^N} \langle x, \mathbf{\Sigma} x \rangle \\
&\langle x, \mathrm{e} \rangle = 1 \\
&x \geq 0
\end{aligned}
$$

Context and History

Why Minimum Variance Portfolio?

Percent Return Over Time of Market vs. Minimum Variance Portfolio for 1000 Largest US Stocks



(Clarke, de Silva & Thorley 2011)

- Minimum Variance Portfolio has usually beaten Market Portfolio (1000 largest US Equities). This is the Low-Volatility anomaly: low beta stocks get a higher than expected return.

Why Long Only Constraints?

- Markowitz actually included long-only constraints in his original paper; these were subsequently dropped from most formulations.
- "imposing the no-short-sales constraints can help even when the constraints do not hold in the population." (Jagannathan, Ma & Zhang 2003)
- No-Shortselling Constraints acts like a shrinkage estimator that prevents massive weights, i.e. exchanging a little bias for a reduction in variance.

Portfolio Theory Origins of Quadratic Programming

- Mean-Variance gives an early example of an algorithm to solve Quadratic Programs- the Critical Line Algorithm.
- Other Algorithms supplanted it; Markowitz was more focused on portfolio theory than mathematical optimization
- Now we have robust and efficient numerical solvers for many real-world quadratic programs

Semi-explicit forms for long-only minimum variance weghts.

Covariance with a factor structure

We consider a $p \times p$ covariance matrix of the form

$$\boldsymbol{\Sigma} = \mathbf{B}\mathbf{V}\mathbf{B}^\top + \boldsymbol{\Delta} \tag{1}$$

with a $p \times q$ full-rank matrix $\mathbf{B}$, a $q \times q$ diagonal matrix $\mathbf{V}$ and a $p \times p$ diagonal matrix $\boldsymbol{\Delta}$. Both $\mathbf{V}$ and $\boldsymbol{\Delta}$ are positive definite.

Without loss of generality, we take $\mathbf{B}^\top \boldsymbol{\Delta}^{-1} \mathrm{e} \geq 0$.

A factor model $Y = \mathbf{B}X + Z$ with $X \in \mathbb{R}^q$ and $Z \in \mathbb{R}^p$ where $\mathbf{Var}(X) = \mathbf{V}$ and $\mathbf{Var}(Z) = \boldsymbol{\Delta}$ while $\mathbf{Cov}(X, Z) = 0$.

Main result.

Let $\mathbf{\Sigma} = \mathbf{B}\mathbf{V}\mathbf{B}^\top + \mathbf{\Delta}$ with $\mathbf{B}^\top \mathbf{\Delta}^{-1} \mathrm{e} \geq 0$ and consider the QP:

$$\text{(QP)} \qquad \begin{aligned} &\min_{x \in \mathbb{R}^N} \langle x, \mathbf{\Sigma} x \rangle \\ &\langle x, \mathrm{e} \rangle = 1 \\ &x \geq 0. \end{aligned}$$

Define $\varphi : \mathbb{R}^q \to \mathbb{R}^q$ by $\varphi(\vartheta) = \mathbf{V}\mathbf{B}^\top \mathbf{\Delta}^{-1} (\mathrm{e} - \mathbf{B}\vartheta)_+$.

Theorem 1. Let x be a solution of problem QP. Then

$$\text{(2)} \qquad x = \frac{w}{\langle w, \mathrm{e} \rangle} \quad \text{where} \quad w = \mathbf{\Delta}^{-1} (\mathrm{e} - \mathbf{B}\theta)_+$$

for $\theta \in \mathbb{R}^q$ the unique fixed point of φ (i.e., $\varphi(\theta) = \theta$).

Sketch of the proof (A).

Change variables ($x \to y^2$) to enforce the nonnegativity.

In the new variables, the QP takes the simpler form

$$\min_{|y^2|=1} \langle y^2, \boldsymbol{\Sigma} y^2 \rangle \tag{3}$$

Show that the KKT conditions for the QP are solved by a fixed point of φ given by $\varphi(\vartheta) = \mathbf{V}\mathbf{B}^\top \boldsymbol{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\vartheta)_+$.

Sketch of the proof (B).

Show that the KKT conditions for the QP are solved by a fixed point of φ given by $\varphi(\vartheta) = \mathbf{V}\mathbf{B}^\top\mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\vartheta)_+$.

Are there several fixed points of φ? Do we have the right one? Does any $\theta = \varphi(\theta)$ give the right portfolio?

Is the fixed point unique?

Sketch of the proof (C).

We prove $\varphi \triangleq \vartheta \to \mathbf{V}\mathbf{B}^\top \boldsymbol{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\vartheta)_+$ has a unique fixed point.

Show $\mathbb{D} = \{\vartheta \in \mathbb{R}^q \,:\, \varphi(\vartheta) \geq 0\}$ is compact and φ is continuous.

The restriction of φ to $\mathbb{D}$ the has a fixed point by Brouwer's fixed point theorem. Easy to see no fixed point exists on $\mathbb{D}^c$.

Furthermore, $\nabla\varphi$ is negative definite on $\mathbb{D}$ since $\mathbf{B}^\top \boldsymbol{\Delta}^{-1}\mathrm{e} \geq 0$. Therefore the fixed point of φ is unique and lies in $\mathbb{D}$.

Alternatively, one can show a one-to-one correspondence between fixed points of φ and solutions of the KKT conditions.

The portfolio as a fixed point.

Let $w = \mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\theta)_+$ be as in Theorem 2 so that $\theta = \psi(\theta)$ and the weights $x = \frac{w}{\langle w, \mathrm{e} \rangle}$ solve the QP problem.

Define $\Psi : \mathbb{R}^p \to \mathbb{R}^p$ by $\Psi(v) = \mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{BVB}^\top v)_+$.

Corollary. The $w = \mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\theta)_+$ is a fixed point of Ψ. i.e.,

$$w = \Psi(w) \tag{4}$$

Unfortunately, the map Ψ is not a contraction and the iterates $v_{k+1} = \Psi(v_k)$ diverge in general starting arbitrarility close to w!

Computing the fixed point.

The iterates $\vartheta_{k+1} = \varphi(\vartheta_k)$ do not converge as the map φ is not a contraction in general. How do we compute the fixed point?

Define $\psi : \mathbb{R}^q \to \mathbb{R}^q$ as $\psi(\vartheta) = \mathbf{A}_\vartheta^{-1} b_\vartheta$ where

$$
\begin{aligned}
\mathbf{A}_\vartheta &= \mathbf{V}^{-1} + \mathbf{B}^\top \mathbf{\Delta}^{-1} \mathbf{diag}(\chi_\vartheta) \mathbf{B} \\
b_\vartheta &= \mathbf{B}^\top \mathbf{\Delta}^{-1} \chi_\theta \\
\chi_\vartheta &= (\mathrm{e} \geq \mathbf{B}\vartheta) \in \{0, 1\}^p
\end{aligned}
\tag{5}
$$

Lemma 1. We have $\theta = \varphi(\theta)$ if and only if $\theta = \psi(\theta)$.

Lemma 2. The iterates $\vartheta_{k+1} = \psi(\vartheta_k)$ converge to the fixed point of ψ (i.e. $\theta = \psi(\theta)$) provided that $\vartheta_0 = 0 \in \mathbb{R}^q$.

Lemma 1. We have $\theta = \varphi(\theta)$ if and only if $\theta = \psi(\theta)$.

The proof follows basic matrix algebra in both directions.

Remark 1. While φ is continuous, the ψ is a discontinous simple function (i.e., it takes on a finite number of values).

Remark 2. Cannot establish convergence of fixed point iterates for ψ via standard tools such as the Banach fixed point theorem (i.e., these rely on continuity and the contraction property).

Lemma 2.

Lemma 2. The iterates $\vartheta_{k+1} = \psi(\vartheta_k)$ converge to the fixed point of ψ (i.e. $\theta = \psi(\theta)$) provided that $\vartheta_0 = 0 \in \mathbb{R}^q$.

- Show that starting from the origin moving in any direction the function is nondecreasing before the fixed point.
- If there is a (strict) decrease then no fixed point can exist, a contradiction. Thus the iterates are increasing, i.e.,

$$\vartheta_{k+1} = \psi(\vartheta_k) \leq \psi(\vartheta_{k+1}) = \vartheta_{k+2}.$$

- Since the function ψ takes on finitely many values, the iterates increase until the fixed point is attained.

Second result.

Let $\mathbf{\Sigma} = \mathbf{B}\mathbf{V}\mathbf{B}^\top + \mathbf{\Delta}$ with $\mathbf{B}^\top \mathbf{\Delta}^{-1}\mathrm{e} \geq 0$ and consider the QP:

$$\text{(QP)} \qquad \begin{aligned} &\min_{x \in \mathbb{R}^N} \langle x, \mathbf{\Sigma} x \rangle \\ &\langle x, \mathrm{e} \rangle = 1 \\ &x \geq 0. \end{aligned}$$

Define $\psi : \mathbb{R}^q \to \mathbb{R}^q$ as $\psi(\vartheta) = \mathbf{A}_\vartheta^{-1} b_\vartheta$ where

$$\text{(6)} \qquad \begin{aligned} \mathbf{A}_\vartheta &= \mathbf{V}^{-1} + \mathbf{B}^\top \mathbf{\Delta}^{-1} \mathbf{diag}(\chi_\vartheta) \mathbf{B} \\ b_\vartheta &= \mathbf{B}^\top \mathbf{\Delta}^{-1} \chi_\theta \\ \chi_\vartheta &= (\mathrm{e} \geq \mathbf{B}\vartheta) \in \{0, 1\}^p. \end{aligned}$$

Define $\psi : \mathbb{R}^q \to \mathbb{R}^q$ as $\psi(\vartheta) = \mathbf{A}_\vartheta^{-1} b_\vartheta$ where

$$(7) \qquad \begin{aligned} \mathbf{A}_\vartheta &= \mathbf{V}^{-1} + \mathbf{B}^\top \boldsymbol{\Delta}^{-1} \mathbf{diag}(\chi_\vartheta)\mathbf{B} \\ b_\vartheta &= \mathbf{B}^\top \boldsymbol{\Delta}^{-1} \chi_\theta \\ \chi_\vartheta &= (\mathrm{e} \geq \mathbf{B}\vartheta) \in \{0,1\}^p \end{aligned}$$

Theorem 2. Let x be a solution of problem QP. Then

$$(8) \qquad x = \frac{w}{\langle w, \mathrm{e} \rangle} \quad \text{where} \quad w = \boldsymbol{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\theta)_+$$

for $\theta \in \mathbb{R}^q$ the unique fixed point of ψ (i.e., $\psi(\theta) = \theta$). Moreover, the iterates $\vartheta_{k+1} = \psi(\vartheta_k)$ converge to θ provided $\vartheta_0 = 0 \in \mathbb{R}^q$.

Algorithms (fixed-point map)

Python implementation of the function ψ.

```
def psi(theta):

    chi = Dinv * (ones >= B @ theta)
    b = B.T @ chi
    A = Vinv + (B.T @ np.diag(chi)) @ B

    return solve(A,b)
```

Algorithms (fixed-point iteration)

Python implementation of the fixed point iterations.

```
def compute_fixed_point (t):

    s = t + 2 * tol

    while (norm(s - t) > tol * norm(s)):
        t = s
        s = psi(t)

    return(s)
```

Algorithms (portfolio computation)

Python computation of the long-only portfolio weights.

```
def lo_weights ():

    theta = psi (np.zeros(q))
    w = Dinv @ (ones - B @ theta)

    return w / sum(w)
```

Comparison to minimum variance

Let $\mathbf{\Sigma} = \mathbf{BVB}^\top + \mathbf{\Delta}$ with $\mathbf{B}^\top \mathbf{\Delta}^{-1} \mathrm{e} \geq 0$ and consider the QP:

$$\begin{aligned} &\min_{x \in \mathbb{R}^N} x^\top \mathbf{\Sigma} x \\ &\langle x, \mathrm{e} \rangle = 1. \end{aligned} \tag{9}$$

The solution is well-known to be $x = \frac{w}{\langle w, \mathrm{e} \rangle}$ where

$$w = \mathbf{\Sigma}^{-1} \mathrm{e}. \tag{10}$$

Applying Woodbury's identity we can compare the two portfolios.

Comparison to minimum variance

LS – Long-short portfolio / LO – Long-only portfolio

Quantity	LS Min. Var.	LO Min. Var.
weights	$x = \frac{w}{\langle w, \mathrm{e} \rangle}$	$x = \frac{w}{\langle w, \mathrm{e} \rangle}$
w	$w = \mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\theta)$	$w = \mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\theta)_+$
θ	$\theta = \varphi(\theta)$	$\theta = \varphi(\theta)$
$\varphi(\vartheta)$	$\mathbf{V}\mathbf{B}^\top\mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\vartheta)$	$\mathbf{V}\mathbf{B}^\top\mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\vartheta)_+$
fixed point	$w = \Psi(w)$	$w = \Psi(w)$
$\Psi(v)$	$\mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\mathbf{V}\mathbf{B}^\top v)$	$\mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\mathbf{V}\mathbf{B}^\top v)_+$

**A similar description holds for the disconinuous map ψ which shares the fixed point with φ but it takes longer to write.*

Interpretation

Our results lead to an understanding of how the parameters $(\mathbf{B}, \mathbf{V}, \mathbf{\Delta})$ that form $\mathbf{\Sigma} = \mathbf{B}\mathbf{V}\mathbf{B}^\top + \mathbf{\Delta}$ influence the QP solutions.

All the solutions are of the form $x = \frac{w}{\langle w, \mathrm{e} \rangle} \in \mathbb{R}^p$ where

$$w = \mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}\theta)_+ \tag{11}$$

and $\theta = \varphi(\theta)$ contains the dependence on $\mathbf{V}$.

- *The dependence on $\mathbf{V}$ vanishes in the problem size p.*
- *Solutions are inversely proportional the diagonals of $\mathbf{\Delta}$.*
- *For the $\beta^k \in \mathbb{R}^p$ the k column (factor) of $\mathbf{B}$, the solutions depend on the linear combination $\theta_1 \beta^1 + \cdots + \theta_q \beta^q$.*

Interpretation

For the $\beta^k \in \mathbb{R}^p$ the k column (factor) of $\mathbf{B}$ we have

$$\mathbf{B}\theta = \theta_1 \beta^1 + \cdots + \theta_q \beta^q \tag{12}$$

is a linear combination of the factors and each θ_k measures the contribution of the kth factor to the portfolio.

The fixed point $(\theta_1, \dots, \theta_q) = \theta = \varphi(\theta)$ thus determines the *theresholds* for the inclusion of each factor.

One-Factor Theorem Analysis

One Factor Model with Long Only Constraints

Let $\beta \in \mathbb{R}^p$, and $\delta_i^2, \sigma^2 \in \mathbb{R}_+$. Under a single covariance factor, our minimization problem now becomes:

$$\begin{aligned} &\min_{x \in \mathbb{R}^p} x^\top (\sigma^2 \beta \beta^\top + \delta_i^2) x \\ &\text{subject to:} \\ &\begin{cases} \sum_k x_k = 1, \\ x_i \geq 0 \end{cases} \end{aligned} \tag{13}$$

Recall, the solution is given by:

$$x_i \propto w_i = \frac{(1 - \theta \beta_i)_+}{\delta_i^2} \tag{14}$$

Note that this is the Sharpe Model/CAPM.

Prior Work on Long-Only Factor Models

- 1-Factor version published in 2011, with some slight caveats (Clarke et al. 2011)
- Explained why large-β_i weights are truncated to 0
- Proof was "semi-formal" and difficult to follow.

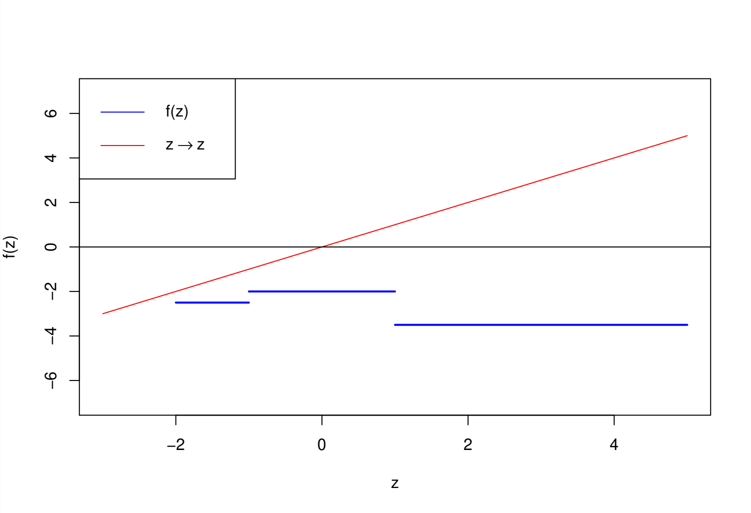
Clarke Formula

- Solution involved computation Fixed Point $\eta = \psi_c(\eta)$ of the awkward function:

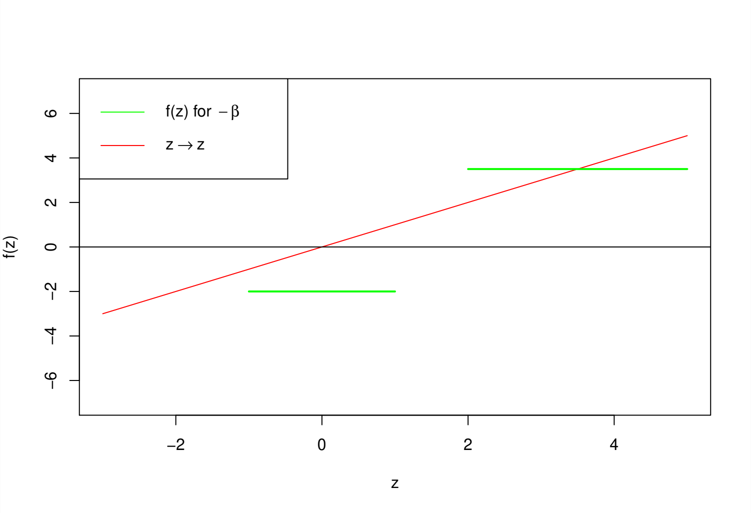
$$\psi_c(z) = \frac{1/\sigma^2 + \sum_{\{i:\beta_i<z\}} \beta_i^2/\delta_i^2}{\sum_{\{i:\beta_i<z\}} \beta_i/\delta_i^2} \tag{15}$$

- Solutions may not exist: e.g. $\beta = [-2, -1, 1]^\top$ and $\sigma^2 = \delta^2 = 1$.
- Instead, of a fixed point, may have $\psi_c(\eta) = -\eta$; alternatively, a fixed point for $-\beta$.

Existence Counter-Example



Existence Counter-Example



One Factor Model with Long Only Constraints (cont.)

Reformulated Clarke Map. First, assume without loss of generality $e^\top \beta > 0$. Define $\psi : \mathbb{R}_+ \mapsto \mathbb{R}$:

$$\psi(z) = \frac{\sum_{\{i : z\beta_i < 1\}} \beta_i / \delta_i^2}{1/\sigma^2 + \sum_{\{i : z\beta_i < 1\}} \beta_i^2 / \delta_i^2} \tag{16}$$

- This is the reciprocal of ψ_C with a slightly different summand condition.
- This is the same as the multidimensional $\psi(z) = \mathbf{A}_z^{-1} b_z$ system introduced earlier, but with $q = 1$.
- We wish to find θ such that $\psi(\theta) = \theta$.
- $\psi(\cdot)$ is still a discontinuous step function; difficult to work with.

Lemma
Assume $\psi(0) > 0$ (equiv. $\mathrm{e}^\top \beta > 0$). Define the function $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$:

$$\phi(z) = -\sigma^2 \left(\sum_{\{i : z\beta_i < 1\}} \frac{\beta_i}{\delta_i^2} \left(z\beta_i - 1 \right) \right). \tag{17}$$

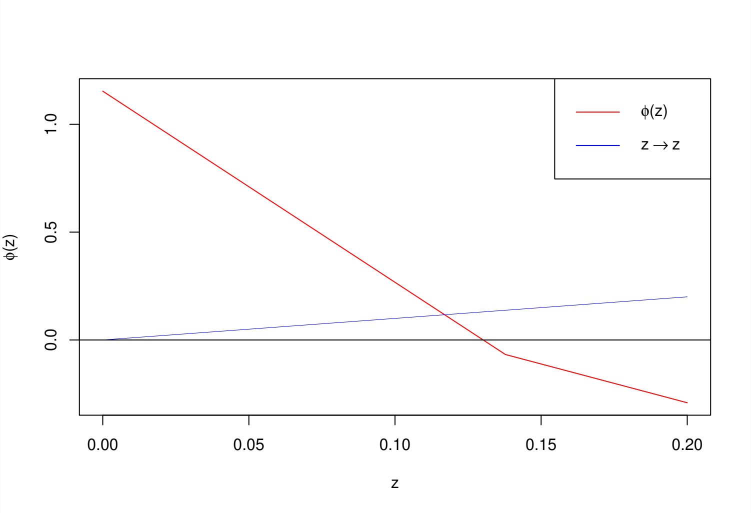
A fixed point $\phi(\theta) = \theta$ exists if and only if a fixed point of $\psi(\theta) = \theta$ exists; furthermore, the solutions have the same value.

Note that is is also the same as the $\varphi(z) = \mathbf{V}\mathbf{B}^\top \mathbf{\Delta}^{-1}(\mathrm{e} - \mathbf{B}z)_+$ function for $q = 1$.

Lemma
For $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$ as defined in Equation (17), we have the following:

1. $\phi(z)$ *is continuous*
2. $\phi(0) > 0$
3. ϕ *is differentiable a.e. and* $\phi'(z) < 0$ *for all* $z > 0$
4. $\phi(z)$ *has a positive root.*

One Factor Model with Long Only Constraints (cont.)



One Factor Model with Long Only Constraints (cont.)

Theorem
There exists a unique nonzero θ that solves $\phi(\theta) = \theta$

This solution is equivalent to the solution to Equation (16).

Proof.
This is a straightforward consequence of the Brouwer Fixed-Point Theorem, as $\phi(\cdot)$ is continuous, $\phi(0) > 0$ and for some $z > 0$, $\phi(z) = 0$. $\square$

Algorithm for One-Factor Model

Goal: Compute the fixed-point in Equation (16) under the assumption that $e^\top \beta/\delta^2 > 0$ and $\theta > 0$.
Note: For θ large enough, $\psi(\theta)$ decreasing and the condition $\theta\beta_i < 1$ ensures that every negative β_i will be included in the summation.

- Order β vector in increasing order.
- Starting from $k = p$, compute

$$\psi_k = \frac{\sum_{i=1}^{k} \beta_i/\delta_i^2}{1/\sigma^2 \sum_{i=1}^{k} \beta_i^2/\delta_i^2} \tag{18}$$

- If $\psi_k\beta_k < 1$, stop, otherwise repeat for $k = k - 1$.

Case Study

Case Study: Sensitivity with Respect to Factor Variance

- We now have semi-explicit formulae for the solutions of Long-Only Markowitz Portfolios which allow us to conduct a sensitivity analyses.
- We want to understand the sensitivity w_i (and thus x_i) with respect to factor variance, σ^2.
- With our formulae, we do not need to re-solve the quadratic program for every value we wish to check

Case Study: Sensitivity with Respect to Factor Variance

For a fixed θ, we can examine the following partial derivatives:

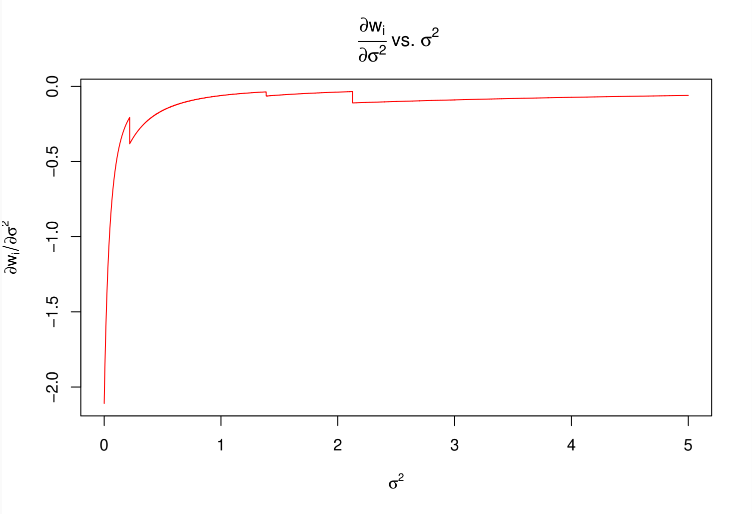
$$\frac{\partial}{\partial \sigma^2} w_i = -\frac{\beta_i}{\delta_i^2} \frac{\partial \theta}{\partial \sigma^2} \tag{19}$$

$$\frac{\partial}{\partial \sigma^2} \theta = \frac{(\phi(\theta)/\sigma^2)}{1 + \sigma^2 \sum_{\{i : \theta \beta_i < 1\}} \beta_i^2 / \delta_i^2} \tag{20}$$

$$\frac{\partial}{\partial \sigma^2} w_i = \frac{\beta_i \sum_{\{k : \theta \beta_k < 1\}} (\beta_k / \delta_k^2)(\theta \beta_k - 1)}{\delta_i^2 + \delta_i^2 \sigma^2 \sum_{\{k : \theta \beta_k < 1\}} \beta_k^2 / \delta_k^2} \tag{21}$$

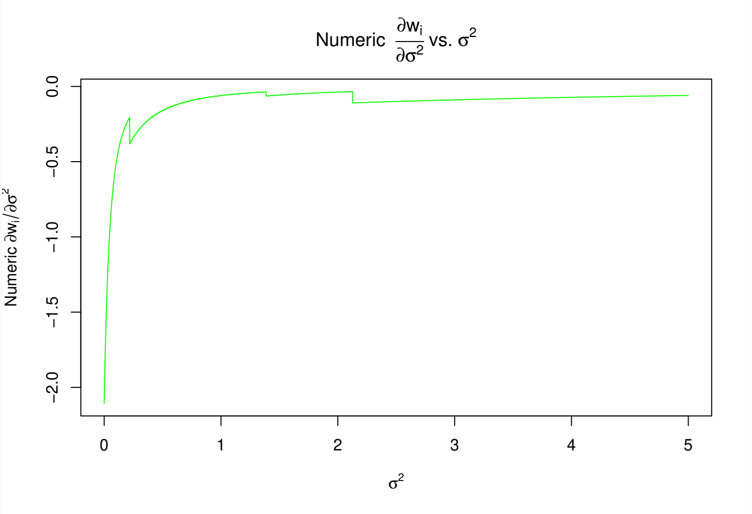
These are ugly formulae...

Derivative of w_i with Respect to σ^2



$p = 5,\ \beta \sim \mathcal{N}_5(1, 1),\ \delta_i^2 \sim \text{lognorm}(0, 1).$

Derivative of w_i with Respect to σ^2



$p = 5$, $\beta \sim \mathcal{N}_5(1, 1)$, $\delta_i^2 \sim \text{lognorm}(0, 1)$.

Numeric Derivative is of θ, computed over many values of σ^2.

Comments

- Numeric derivative computed by differencing solutions of θ rather than re-solving the quadratic program.
- Numeric derivative and actual formula are very close
- Furthermore, these can be very computationally difficult to estimate even in the 1-Factor case without Fixed-Point formulation

Future Work

- Include risk tolerance, and map these results to the entire efficient frontier
- Generalize long-only constraints to arbitrary weight constraints
- Study the implications of these weights for statistical estimation of covariance matrices

Clarke, Roger, Harindra de Silva & Steven Thorley (2011), 'Minimum-variance portfolio composition', *The Journal of Portfolio Management* **37**(2), 31–45.
URL: *https://jpm.iijournals.com/content/37/2/31*

Jagannathan, Ravi, Tongshu Ma & Jiaqi Zhang (2003), 'A note on "risk reduction in large portfolios: Why imposing the wrong constraints helps"', *The Journal of Finance* **74**(5), 2689–2696.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12824*