Parametric Linear Programming and Portfolio Optimization

Robert J. Vanderbei

(joint with A. Ruszczyński)

2007 March 20

Cowles Foundation for Research in Economics
Workshop on Optimization
Yale University

Operations Research and Financial Engineering
Princeton University
Princeton, NJ 08544
http://www.princeton.edu/∼rvdb

ABSTRACT

1. The traditional (quadratic) Markowitz model produces portfolios that are *stochastically dominated* by portfolios not on the efficient frontier. This is BAD.
2. Replacing the quadratic risk measure with a mean absolute deviation (MAD) measure corrects this defect.
3. The MAD model can be formulated as a parametric linear programming problem (the risk parameter λ is the parameter).
4. The *parametric* simplex method can be used with λ as the parameter of the parametric method.
5. Doing so, one finds ALL portfolios on the efficient frontier in roughly the same time as it takes to find just one portfolio (corresponding, say, to $\lambda = 0$).
6. The speedup is huge.
7. The parametric simplex method has other useful features—to be discussed time permitting.

Markowitz Shares the 1990 Nobel Prize

Press Release - The Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel

KUNGL. VETENSKAPSAKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

16 October 1990

THIS YEAR'S LAUREATES ARE PIONEERS IN THE THEORY OF FINANCIAL ECONOMICS AND CORPORATE FINANCE

The Royal Swedish Academy of Sciences has decided to award the 1990 Alfred Nobel Memorial Prize in Economic Sciences with one third each, to

Professor **Harry Markowitz**, City University of New York, USA,
Professor **Merton Miller**, University of Chicago, USA,
Professor **William Sharpe**, Stanford University, USA,

for their pioneering work in the theory of financial economics.

Harry Markowitz is awarded the Prize for having developed the theory of portfolio choice;
William Sharpe, for his contributions to the theory of price formation for financial assets, the so-called, *Capital Asset Pricing Model* (CAPM); and
Merton Miller, for his fundamental contributions to the theory of corporate finance.

Summary
Financial markets serve a key purpose in a modern market economy by allocating productive resources among various areas of production. It is to a large extent through financial markets that saving in different sectors of the economy is transferred to firms for investments in buildings and machines. Financial markets also reflect firms' expected prospects and risks, which implies that risks can be spread and that savers and investors can acquire valuable information for their investment decisions.

The first pioneering contribution in the field of financial economics was made in the 1950s by Harry Markowitz who developed a theory for households' and firms' allocation of financial assets under uncertainty, the so-called theory of portfolio choice. This theory analyzes how wealth can be optimally invested in assets which differ in regard to their expected return and risk, and thereby also how risks can be reduced.

Hedging

Investment A: up 20%, down 10%, equally likely—a risky asset.

Investment B: up 20%, down 10%, equally likely—another risky asset.

Correlation: up years for A are down years for B and vice versa.

Portfolio—half in A, half in B: up 5% every year! No risk!

The Ingredients: Risk and Reward

Raw Data:

$$R_j(t) \;=\; \text{return on asset } j \text{ in time period } t$$

Note: R_j is a random variable with the t's forming the sample space

Derived Data:

$$\mu_j = \frac{1}{T}\sum_{t=1}^{T} R_j(t) = \mathbb{E}R_j$$

$$D_{tj} = R_j(t) - \mu_j.$$

Decision Variables:

$$x_j \;=\; \text{fraction of portfolio to invest in asset } j$$

$$R(x) = \sum_j x_j R_j$$

Decision Criteria:

$$\mu(x) = \sum_j \mu_j x_j$$

$$\rho_{\mathsf{var}}(x) = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_j D_{tj}x_j\right)^2 = \mathsf{Var}(R(x))$$

$$\rho_{\mathsf{mad}}(x) = \frac{1}{T}\sum_{t=1}^{T}\left|\sum_j D_{tj}x_j\right|$$

$$\rho_p(x) = \mathbb{E}\left[\frac{1-p}{p}(q_p(x) - R(x)) \vee (R(x) - q_p(x))\right]$$

Quadratic Markowitz Problem

$$\text{maximize} \quad \lambda \sum_j \mu_j x_j - \frac{1}{T} \sum_{t=1}^{T} \left(\sum_j D_{tj} x_j \right)^2$$

$$\text{subject to} \quad \sum_j x_j = 1$$

$$x_j \geq 0 \qquad \text{for all investments } j$$

λ is the risk parameter.

MAD Markowitz Problem

$$\begin{array}{lll} \text{maximize} & \lambda \sum_j \mu_j x_j - \frac{1}{T} \sum_{t=1}^{T} \left| \sum_j D_{tj} x_j \right| & \\ \text{subject to} & \sum_j x_j = 1 & \\ & x_j \geq 0 & \text{for all investments } j \end{array}$$

Not a linear programming problem. But it's easy to convert.

There are two reasons why Quadratic Markowitz is bad, whereas MAD is good:

- Variance is a bad risk measure.
- Linear programming (especially parametric LP) is easier/faster than QP.

cV@R Markowitz Problem

$$\begin{array}{lll} \text{maximize} & \lambda\mu(x) - \rho_p(x) & \\ \text{subject to} & \displaystyle\sum_j x_j = 1 & \\ & x_j \geq 0 & \text{for all investments } j \end{array}$$

Not a linear programming problem. But it's easy to convert.

cV@R is also good.

Good Risk Measures

Stochastic Dominance

Second order stochastic dominance characterizes those random variables that every risk averse decision maker would prefer to a given random variable:

Definition *Random variable V (second-order) stochastically dominates random variable S ($V \succeq_2 S$) if and only if $\mathbb{E}(U(V)) \geq \mathbb{E}(U(S))$ for every increasing concave function $U(\cdot)$.*

Theorem *$V \succeq_2 S$ if and only if $F_V^{(2)} \leq F_S^{(2)}$, where*

$$F_V^{(2)}(z) = \int_{-\infty}^{z} \mathbb{P}(V \leq r) dr.$$

Theorem *There are optimal solutions to the quadratic Markowitz model that are stochastically dominated by other (non-optimal) portfolios.*

Theorem *In the MAD Markowitz model, for $\lambda \geq 2$, optimal portfolios are not stochastically dominated.*

Theorem *In the cV@R Markowitz model, for $\lambda \geq 1$, optimal portfolios are not stochastically dominated.*

Proof Outline for Last Theorem

$$R(x) \succeq_2 R(y) \Longrightarrow \mu(x) \geq \mu(y) \tag{1}$$

As with medians, quantiles can be found by optimization:

$$\rho_p(x) = \min_z \mathbb{E}\left(\frac{1-p}{p}(z - R(x)) \vee (R(x) - z)\right)$$

Consider

$$\begin{aligned} G_{R(x)}(p) &:= p\mu(x) - p\rho_p(x) \\ &= \sup_z \left(p\mu(x) - \mathbb{E}(1-p)(z - R(x)) \vee p(R(x) - z)\right) \\ &= \sup_z \left(pz - F^{(2)}_{R(x)}(z)\right) \end{aligned}$$

$$R(x) \succeq_2 R(y) \quad \Longleftrightarrow \quad F^{(2)}_{R(x)} \leq F^{(2)}_{R(y)} \quad \Longrightarrow \quad G_{R(x)}(p) \geq G_{R(y)}(p) \quad \forall p$$

From this last inequality we get

$$\mu(x) - \rho_p(x) \geq \mu(y) - \rho_p(y) \tag{2}$$

The result follows immediately from (1) and (2).

MAD Markowitz: LP Formulation

$$
\begin{array}{lll}
\text{maximize} & \displaystyle \lambda \sum_j \mu_j x_j - \frac{1}{T} \sum_{t=1}^{T} y_t & \\
\text{subject to} & \displaystyle -y_t \leq \sum_j D_{tj} x_j \leq y_t & \text{for all times } t \\
& \displaystyle \sum_j x_j = 1 & \\
& x_j \geq 0 & \text{for all investments } j \\
& y_t \geq 0 & \text{for all times } t
\end{array}
$$

This is a family of LPs parametrized by λ.

Interior point methods could solve a single instance quickly.

But, the parametric simplex method can solve the entire family (over *all* λ, not just some discrete subset) in one fell swoop.

Adding Slack Variables w_t^+ and w_t^-

$$\begin{array}{lrl}
\text{maximize} & \lambda \sum_j \mu_j x_j - \frac{1}{T} \sum_{t=1}^{T} y_t & \\
\text{subject to} & -y_t - \sum_j D_{tj} x_j + w_t^- = 0 & \text{for all times } t \\
 & -y_t + \sum_j D_{tj} x_j + w_t^+ = 0 & \text{for all times } t \\
 & \sum_j x_j = 1 & \\
 & x_j \geq 0 & \text{for all investments } j \\
 & y_t, w_t^-, w_t^+ \geq 0 & \text{for all times } t
\end{array}$$

The Solution for Large λ

Varying the risk bound $0 \leq \lambda < \infty$ produces the *efficient frontier*.

Large values of λ favor reward whereas small values favor minimizing risk.

Beyond some finite threshold value for λ, the optimal solution will be a portfolio consisting of just one asset—the asset j^* with the largest average return:

$$\mu_{j^*} \geq \mu_j \qquad \text{for all } j.$$

It's easy to identify basic vs. nonbasic variables:

- Variable x_{j^*} is basic whereas the remaining x_j's are nonbasic.
- All of the y_t's are basic.
- If $D_{tj^*} > 0$, then w_t^- is basic and w_t^+ is nonbasic. Otherwise, it is switched.

The Basic Optimal Solution for Large λ

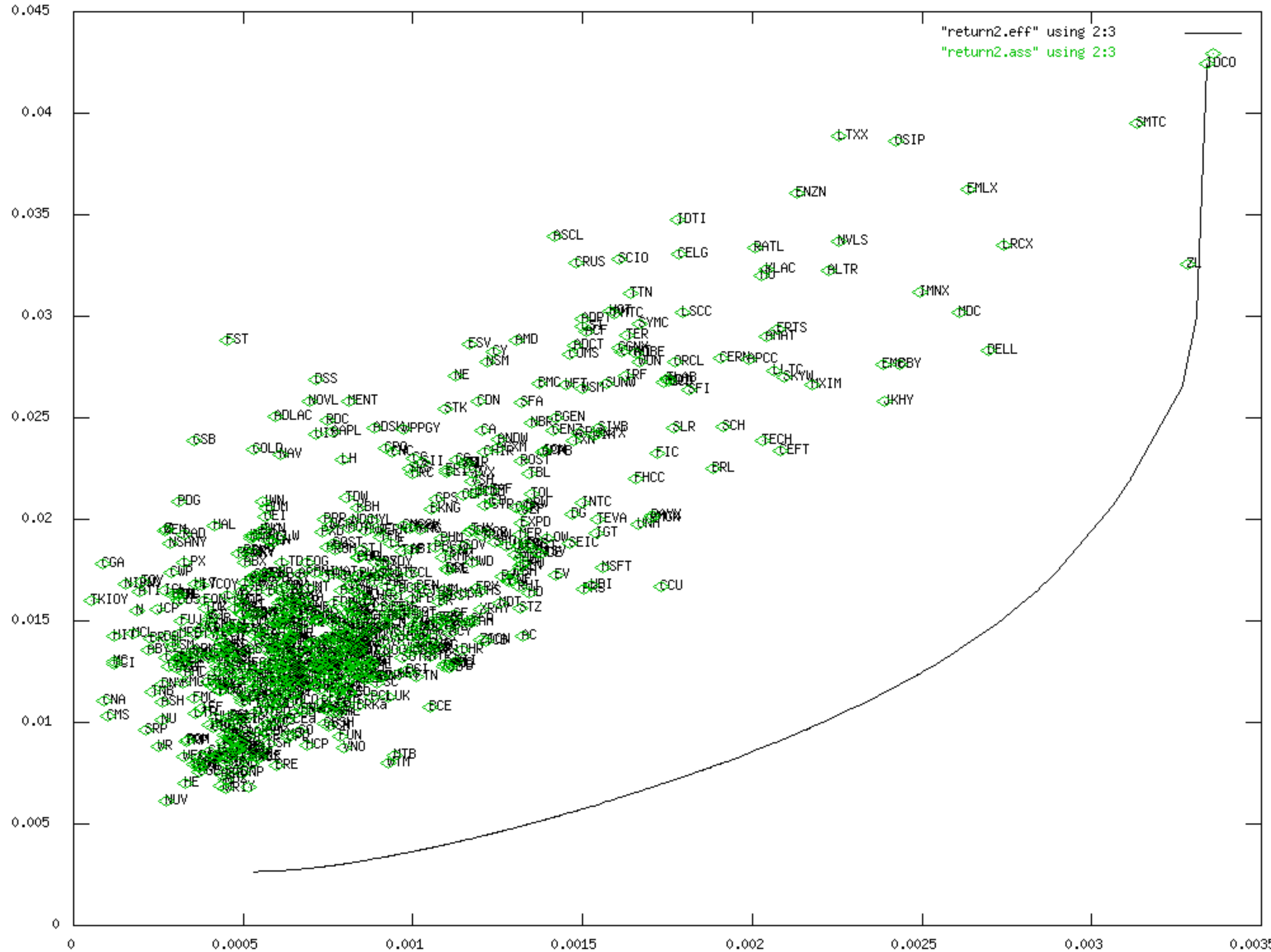
Let

$$T^{+} = \left\{ t : D_{tj^*} > 0 \right\}, \qquad T^{-} = \left\{ t : D_{tj^*} < 0 \right\}, \qquad \text{and} \qquad \epsilon_t = \begin{cases} 1, & \text{for } t \in T^{+} \\ -1, & \text{for } t \in T^{-} \end{cases}$$

It's tedious, but here's the optimal dictionary (for λ large):

$$\begin{array}{lcllll}
\zeta & = & \frac{1}{T}\sum_{t=1}^{T} \epsilon_t D_{tj^*} & -\frac{1}{T}\sum_{j \neq j^*}\sum_{t=1}^{T} \epsilon_t (D_{tj} - D_{tj^*}) x_j & -\frac{1}{T}\sum_{t \in T^-} w_t^- & -\frac{1}{T}\sum_{t \in T^+} w_t^+ \\
 & & +\lambda \mu_{j^*} & +\lambda \sum_{j \neq j^*} (\mu_j - \mu_{j^*}) x_j & & \\
\hline
y_t & = & -D_{tj^*} & -\sum_{j \neq j^*} (D_{tj} - D_{tj^*}) x_j & +w_t^- & & t \in T^- \\
w_t^- & = & 2D_{tj^*} & +2\sum_{j \neq j^*} (D_{tj} - D_{tj^*}) x_j & & +w_t^+ & t \in T^+ \\
y_t & = & D_{tj^*} & +\sum_{j \neq j^*} (D_{tj} - D_{tj^*}) x_j & & +w_t^+ & t \in T^+ \\
w_t^+ & = & -2D_{tj^*} & -2\sum_{j \neq j^*} (D_{tj} - D_{tj^*}) x_j & +w_t^- & & t \in T^- \\
x_{j^*} & = & 1 & -\sum_{j \neq j^*} x_j & & &
\end{array}$$

Daily Returns for 12 Years on 719 Assets



Click here for an expanded browser view.

Computing the Efficient Frontier

Using a reasonably efficient code for the *parametric simplex method* (simpo), it took *22,000* pivots and *1.5 hours* to solve for *one point* on the efficient frontier.

Customizing this same code to solve parametrically for every point on the efficient frontier, it took *23,446* pivots and *57 minutes* to compute *every point* on the frontier.

The efficient frontier consists of *23,446* distinct portfolios. Click here for a partial list (*warning: the file is 2.5 MBytes*). The complete list makes a 37 MByte file.

Description of the Parametric Simplex Method

General Problem:

$$\begin{aligned} \text{maximize } \zeta(x) &= c^T x \\ \text{subject to: } Ax &= b \\ x &\geq 0. \end{aligned}$$

Identify partition of variables into *basic* and *nonbasic*

$$x \stackrel{\mathrm{R}}{=} \begin{bmatrix} x_{\mathcal{B}} \\ x_{\mathcal{N}} \end{bmatrix}$$

Similarly, rearrange columns of A and rows of c:

$$A \stackrel{\mathrm{R}}{=} \begin{bmatrix} B & N \end{bmatrix} \qquad c \stackrel{\mathrm{R}}{=} \begin{bmatrix} c_{\mathcal{B}} \\ c_{\mathcal{N}} \end{bmatrix}$$

Dictionary (or tableau) arrangement:

$$\begin{aligned} \zeta &= \bar{\zeta} - \bar{z}_{\mathcal{N}}^T x_{\mathcal{N}} \\ x_{\mathcal{B}} &= \bar{x}_{\mathcal{B}} - B^{-1} N x_{\mathcal{N}}, \end{aligned}$$

where

$$\bar{x}_{\mathcal{B}} = B^{-1} b, \qquad \bar{z}_{\mathcal{N}} = (B^{-1}N)^T c_{\mathcal{B}} - c_{\mathcal{N}}, \qquad \bar{\zeta} = c_{\mathcal{B}}^T B^{-1} b$$

Dictionary solution:

$$x_{\mathcal{N}} = 0 \qquad x_{\mathcal{B}} = \bar{x}_{\mathcal{B}}$$

Optimal iff:

$$\text{Primal Feasible } (\bar{x}_{\mathcal{B}} \geq 0) \qquad \text{and} \qquad \text{Dual Feasible } (\bar{z}_{\mathcal{B}} \geq 0)$$

Parametric perturbation:

$$\begin{array}{rclccc}
\zeta & = & \bar{\zeta} & +l\lambda & -\bar{z}_{\mathcal{N}}^T x_{\mathcal{N}} \\
 & & +l\lambda & +q\lambda^2 & -\lambda \hat{z}_{\mathcal{N}}^T x_{\mathcal{N}} \\
x_{\mathcal{B}} & = & \bar{x}_{\mathcal{B}} & +\lambda \hat{x}_{\mathcal{B}} & -B^{-1} N x_{\mathcal{N}}.
\end{array}$$

Initially pick $\hat{x}_{\mathcal{B}} > 0$ and $\hat{z}_{\mathcal{N}} > 0$ (and, of course, $l = 0$ and $q = 0$) so that dictionary solution is optimal for λ large.

The algorithm is a homotopy method in which λ is driven from ∞ to 0 all the while maintaining optimality of the current dictionary.

Optimality is maintained by doing standard primal/dual simplex pivots.

Two Reasons to Love the Parametric Simplex Method

- Randomize the coefficients defining the perturbation according to a probability distribution for which $\mathbb{P}(C_1 = C_2) = 0$ whenever C_1 and C_2 are independent random variables having this distribution. For example, uniform on $[0, 1]$ works. Then, for $\lambda \neq 0$, the dictionary will be nondegenerate with probability one.

- Assuming, no degenerate pivots, a simple thought experiment suggests that the expected number of pivots is $(n + m)/2$. Experimental tests support this conjecture.

- A portfolio is *bad* if another portfolio dominates it (stochastically).
- Many portfolios on Markowitz's "efficient frontier" are bad.
- MAD Markowitz isn't bad.
- MAD Markowitz is a parametric LP.
- Even more, using the parametric simplex method the entire efficient frontier can be computed in the time normally required to find just one point on the frontier.
- Lastly, our efficient frontier is completely determined by a finite set of portfolios (vs. a continuum).

References

[1] A. Ruszczyński and R.J. Vanderbei. Frontiers of Stochastically Nondominated Portfolios. *Econometrica*, 71(4):1287–1297, 2003.

Michael Yang, CIO

Steven Kamara, Trade Associate

Currently seeking advice (from accredited investors only) to establish active investment management track record with Preceptor Funds, LLP.

Long-biased risk-managed stock-fund strategy

See Michael or Steven for more information. www.preceptorfund.com

Contents