PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The General Mean-Variance Portfolio Selection Problem [and Discussion]

Harry M. Markowitz, R. Lacey, J. Plymen, M. A. H. Dempster and R. G. Tompkins

Phil. Trans. R. Soc. Lond. A 1994 **347**, 543-549
doi: 10.1098/rsta.1994.0063

The general mean-variance portfolio selection problem

By Harry M. Markowitz
1010 Turquoise Street, Suite 245, San Diego, California 92109, U.S.A.

This paper states the 'general mean-variance portfolio analysis problem' and its solution, and briefly discusses its use in practice.

1. The problem

We consider n securities whose returns $r' = (r_1, \ldots, r_n)$ during the forthcoming period have expected values $\mu' = (\mu_1, \ldots, \mu_n)$ and a covariance matrix $C = (\sigma_{ij})$. An investor is to select a portfolio $X' = (X_1, \ldots, X_n)$. The return $R = r'X$ on the portfolio has expected value and variance, respectively,

$$E = \mu' X, \quad V = X'CX. \tag{1a, b}$$

The portfolio is to chosen subject to constraints

$$AX = b, \quad X \geqslant 0, \tag{1c, d}$$

where A is $m \times n$ and b is $m \times 1$. Thus, non-negative X_i are to be chosen subject to $m \geqslant 1$ linear inequalities.

A portfolio is *feasible* if it satisfies (1c) and (1d). An EV combination is feasible if it is the E and V of a feasible portfolio. A feasible EV combination (E_0, V_0) is *inefficient* if there is another feasible EV combination (E_1, V_1) such that either

$$\text{(i)}\ E_1 > E_0 \ \text{ and } \ V_1 \leqslant V_0$$

or

$$\text{(ii)}\ V_1 < V_0 \ \text{ and } \ E_1 \geqslant E_0.$$

A feasible EV combination is efficient if it is not inefficient. A feasible portfolio is efficient or inefficient in accordance with its EV combination.

It is not sufficient to require only condition (i) or condition (ii) in the definition of inefficiency, nor to define an efficient EV combination as one which maximizes E for given V and minimizes V for given E. Examples can be constructed of feasible EV combinations which meet the latter two requirements but are nevertheless inefficient as previously defined.

Since C is a covariance matrix it is positive semi-definite. We do not require it to be positive definite because C is singular in some important applications. We will see an example below. If C is singular, there may be more than one efficient portfolio which has a given efficient EV combination. We define a 'complete, non-redundant' set of efficient portfolios as one which contains one and only one efficient portfolio for each efficient EV combination.

The portfolio analysis problem is to determine

1. whether constraints (1c) and (1d) are feasible and, if they are, calculate
2. the set of all efficient EV combinations and
3. a complete, non-redundant set of efficient portfolios.

Phil. Trans. R. Soc. Lond. A (1994) **347**, 543–549
Printed in Great Britain

Examples exist of portfolio selection problems with feasible portfolios but no efficient portfolios. This is possible (but not necessary) when E is unbounded and C singular. Excluding this case, every feasible portfolio selection problem has a piecewise parabolic set of efficient EV combinations, with a finite number of pieces. For every such problem there exists a piecewise linear complete, non-redundant set of efficient portfolios. For these assertions to apply with the generality stated, we define 'piecewise' to include a single 'piece' (line or parabolic segment) or only a point. One piece (line segment and parabolic segment) may be unbounded in one direction.

We will consider the computation of these efficient sets after we further consider problem definition and application.

2. Application

It might seem that we could gain some generality by allowing linear inequalities ($\geqslant$, $\leqslant$) in (1c) or allowing some or all variables to be negative. This is not the case. Specifically, given any mean-variance portfolio analysis problem whose constraints permit some or all variables to be negative and permits some or all linear constraints in (1c) to be (weak) inequalities, there is a problem in the standard form of (1c) and (1d) that has the same answer. That is, it has the same set of efficient EV combinations (even the same set of feasible EV combinations) and, given the complete, non-redundant set of efficient portfolios for the equivalent problem it is easy to determine this for the original problem. (See chapter 2 of Markowitz (1987) for details.)

Thus, as far as applications are concerned, we think of the general mean-variance portfolio selection problem as one of finding mean-variance efficient portfolios in variables that may or may not be required to be non-negative, which are subject to one or more (actually, zero or more) linear equalities or weak inequalities. (Also, certain nonlinear constraints can be approximated, as in linear programming.)

Examples of such linear constraints are the budget constraint, a 'turnover constraint' which limits the amount by which the new portfolio may differ from the previous one, and upper or lower bounds on the amount invested in a security or an industry. A special case is an 'exogenous asset' whose amount is fixed in the portfolio. The exogenous asset may, for example, be a random source of income other than the return on securities. For a given E, the variance minimizing portfolio as a whole, including the exogenous asset, depends on the covariances between the exogenous asset and the securities whose amounts are to be selected.

The exogenous asset may be a state variable for which the investor wishes to seek or avoid correlation of portfolio return. The single period mean-variance analysis should be thought of as an approximation to the single period derived utility maximization which is optimal within a many-period investment game. If the derived utility function depends on state variables other than end-of-period wealth, the mean variance approximation may use exogenous assets. (See chapter 3 of Markowitz (1987), starting from the section 'Why mean and variance'.)

We noted that one can convert a model with inequality constraints to an equivalent one with equality constraints. This involves introducing slack variables, as in linear programming. Since slack variables have zero variance, their presence makes C singular. Singular C is no problem for the 'critical line' algorithm described in the next section.

3. Computation

Every feasible general portfolio selection problem has a solution of the following nature. Each piece of the piecewise linear set of efficient portfolios has a set $\mathrm{IN} \subset \{1, \ldots, n\}$ of 'in' securities. The others are 'out'. Obtain A_{IN} and μ_{IN} by setting $a_{ki} = 0$ and $\mu_i = 0$ if $i \in \mathrm{OUT}$. Obtain C_{IN} by setting $\sigma_{ij} = \delta_{ij}$ ($= 1$ if $i = j$, otherwise $= 0$) if either i or j is out.

If

$$M_{\mathrm{IN}} = \begin{pmatrix} C_{\mathrm{IN}} & A'_{\mathrm{IN}} \\ A_{\mathrm{IN}} & 0 \end{pmatrix} \tag{2a}$$

is non-singular, then the solution to

$$M_{\mathrm{IN}} = \begin{pmatrix} X \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix} + \begin{pmatrix} \mu_{\mathrm{IN}} \\ 0 \end{pmatrix} \lambda_E, \tag{2b}$$

i.e.

$$\begin{pmatrix} X \\ \lambda \end{pmatrix} = M_{\mathrm{IN}}^{-1} \begin{pmatrix} 0 \\ b \end{pmatrix} + M_{\mathrm{IN}}^{-1} \begin{pmatrix} \mu_{\mathrm{IN}} \\ 0 \end{pmatrix} \lambda_E = \alpha_{\mathrm{IN}} + \beta_{\mathrm{IN}} \lambda_E, \tag{2c}$$

is referred to as a 'critical line'.

Typically, most critical lines contain no efficient portfolios. If an efficient point exists on a critical line, then there is an interval of efficient portfolios, namely, the solution to (2) for

$$\lambda_E \in [\lambda_{\mathrm{LOW}}, \infty) \quad \text{or} \quad \lambda_E \in [\lambda_{\mathrm{LOW}}, \lambda_{\mathrm{HI}}]. \tag{3a, b}$$

This is referred to as the 'efficient segment' of the critical line.

In particular, there exists a complete, non-redundant set of efficient portfolios which consists of the efficient segments of $K \geqslant 1$ critical lines. These critical lines are efficient for

$$\lambda_E \in [\lambda^1_{\mathrm{LOW}}, \infty), \quad \lambda_E \in [\lambda^2_{\mathrm{LOW}}, \lambda^2_{\mathrm{HI}}], \ldots, \lambda_E \in [0, \lambda^K_{\mathrm{HI}}],$$

where $\lambda^{k+1}_{\mathrm{HI}} = \lambda^k_{\mathrm{LOW}}$ and $\lambda^k_{\mathrm{HI}} \geqslant \lambda^k_{\mathrm{LOW}}$.

IN^{k+1} differs from IN^k by the addition or deletion of one member. In the relatively easy to explain 'non-degenerate' case, to be defined below, $\lambda^k_{\mathrm{HI}} > \lambda^k_{\mathrm{LOW}}$ for $k = 1, \ldots, K$.

Define $\eta' = (\eta_1, \ldots, \eta_n)$ by

$$\eta = (C_{\mathrm{IN}}, A'_{\mathrm{IN}}, -\mu_{\mathrm{IN}}) \begin{pmatrix} X \\ \lambda \\ \lambda_E \end{pmatrix}. \tag{4}$$

η_i is the partial derivative of a lagrangian with respect to X_i. Substituting (2c) into (4) we obtain η as a linear function of λ_E

$$\eta = \gamma_{\mathrm{IN}} + \delta_{\mathrm{IN}} \lambda_E. \tag{5}$$

A sufficient condition for a point on a critical line to be an efficient portfolio is that

$$X \geqslant 0, \quad \eta \geqslant 0, \quad \lambda_E > 0. \tag{6a–c}$$

A portfolio satisfying (6a), (6b) and $\lambda_E = 0$ may or may not be efficient. We return to this point below.

The critical line algorithm computes a complete, non-redundant set of efficient portfolios. The computation is simplest when the problem is non-degenerate (still to be defined) and feasible E is bounded above. Part of the non-degeneracy assumptions (to be relaxed below) are that (a) there is a unique feasible portfolio which maximizes E, and (b) the problem of maximizing E subject to (1c) and (1d) is non-degenerate in the sense defined in linear programming. In this case, the unique optimum solution has exactly m variables – the 'basis variables' – with $X_i > 0$. These basis variables are the first IN set. With IN_1 thus defined, M_{IN} is non-singular, and M_{IN}^{-1} is easy to compute. The critical line

$$\begin{pmatrix} X \\ \lambda \end{pmatrix} = \alpha + \beta\lambda_E, \quad \eta = \gamma + \delta\lambda_E, \tag{7a, b}$$

satisfies (6a, b) and $\lambda_E \geqslant 0$ for all

$$\lambda_E \in [\lambda^1_{\mathrm{LOW}}, \infty),$$

where λ_{LOW} is the largest λ_E below which one of the three conditions becomes false; that is, it is the first (i.e. largest) λ_E at which $\eta_i \downarrow 0$ for some OUT i or $\lambda_E \downarrow 0$. With subsequent IN sets we also have the possibility that $X_i \downarrow 0$ for some IN i; but with IN_1, i.e. on the first critical line, X is constant and equal to the E maximizing portfolio. Whereas (6a, b) and $\lambda_E = 0$ do not in general assure an efficient portfolio, if $\lambda^1_{\mathrm{LOW}} = 0$ then the portfolio with maximum E is also the efficient portfolio with minimum V, and is thus the entire complete non-degenerate set of efficient portfolios. We next consider the case of $\lambda^1_{\mathrm{LOW}} > 0$.

The remaining non-degeneracy assumption is that, at each iteration k, a unique i determines $\lambda^k_{\mathrm{LOW}} > 0$. In particular, in the first iteration a unique $\eta_{i*} \downarrow 0$ first. IN_2 is the same as IN_1 except for the addition of $i*$. The new M_{IN} associated with IN_2 is guaranteed to be non-singular.

The fine print in this guarantee says that the statement would be true if we performed calculations with unlimited precision. This could actually be done by storing the numerator and denominator of all results as unlimited, but always finite, integers; since only rational operations are performed. In practice, floating point arithmetic is used, round-off errors occur, and the aforementioned and other 'guarantees' are no longer certain. As a matter of fact, the computation succeeds, including satisfying conditions which assure optimality, most of the time even for $n \geqslant 1000$. When it fails, some rescaling or a practically equivalent restatement of the problem usually, perhaps always, succeeds. The fact that the algorithm usually works does not diminish the desirability of understanding the round-off error problem much better than we do. From this point on we ignore the round-off error problem and, in effect, assume that computations are performed with unlimited precision.

Given our current non-degeneracy assumptions, X_{i*} will increase as λ_E is reduced below λ^1_{LOW}. Conditions (6a, b) and $\lambda_E \geqslant 0$ will hold for

$$\lambda_E \in [\lambda^2_{\mathrm{LOW}}, \lambda^2_{\mathrm{HI}}],$$

where $\lambda^2_{\mathrm{HI}} = \lambda^1_{\mathrm{LOW}}$, $\lambda^2_{\mathrm{LOW}} < \lambda^2_{\mathrm{HI}}$ and $\lambda^2_{\mathrm{LOW}} \geqslant 0$. λ^2_{LOW} is the largest value of $\lambda_E < \lambda^2_{\mathrm{HI}}$ below which one of the three conditions will be violated. If $\lambda^2_{\mathrm{LOW}} = 0$ then the point on the second critical line at $\lambda_E = 0$ is the efficient portfolio with minimum variance. (When C is singular, there may be other feasible portfolios with the same V but lower E.)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

In case $\lambda^2_{\text{LOW}} > 0$, one of our above-stated non-degeneracy assumptions assures us that one and only one i IN will satisfy

$$\alpha_i + \beta_i \lambda^2_{\text{LOW}} = 0, \tag{8a}$$

or i OUT will satisfy

$$\gamma_i + \delta_i \lambda^2_{\text{LOW}} = 0. \tag{8b}$$

($X_i = 0$ for i OUT and $\eta_i = 0$ for i IN at all points on the critical line.)

IN_3 differs from IN_2 by the deletion of the i^* that satisfies (8a) or the addition of the i^* that satisfies (8b). In either case, the new M_{IN} will be non-singular (even if C is singular). In case (8a), η_i will increase as λ_E is reduced below λ^2_{LOW}; in case (8b), X_i will increase as λ_E decreases.

Our discussion of IN_2, its corresponding critical line and their relation to IN_3 and its critical line, illustrates the general case. The same relations hold for IN_k and between IN_k and IN_{k+1}. The computation stops when $\lambda_E = 0$ is reached. This must happen in a finite number of steps, since the same IN set cannot appear twice.

4. Degeneracy and other problems

Now we tie up loose ends. First, we note that certain difficulties are handled for us when we use George Dantzig's simplex algorithm for solving the linear programming problem of maximizing E subject to (1c) and (1d). (See Dantzig (1963) for details concerning the simplex algorithm, cycling in linear programming, etc.) If the model is unfeasible, this is determined by 'phase I' of the simplex calculation, and we are so advised. If the rank of A is less than m, phase I provides an equivalent model which is not rank deficient.

Phase II informs us if feasible E is unbounded or if the E maximizing solution is not unique or is degenerate in the sense that some variable in the basis has $X_i = 0$. It also informs us as to which non-basic activities (columns of A) have 'zero profitability', i.e. have the partial derivative of the appropriate lagrangian equal to zero.

It has been shown that degenerate linear programming problems can cycle; that is, a sequence of iterations can occur in which variables enter and leave the basis but do not change value, and in which a given basis repeats itself. If the same (non-random) rule is followed to decide which non-basic variable with positive profit is to go into the basis, and which basic variable with zero value is to go out, then the once repeating basis will repeat infinitely often.

The problem of modifying the simplex algorithm so it is guaranteed not to cycle can be solved along the following lines. In principle, for sufficiently small positive ϵ, adding certain powers of ϵ to each b_i will produce a still feasible linear program which is not degenerate and has almost the same answer as the original linear program. The answer to this perturbed problem approaches the original answer as $\epsilon \downarrow 0$. It is not necessary to actually add these powers of ϵ, since one can calculate the sequence of bases which would occur for any sufficiently small positive ϵ.

In practice, the simplex algorithm never (or perhaps hardly ever) cycles, except in problems especially constructed to show that cycling is possible. In other words, if one ignores the existence of cycling and arbitrarily breaks ties for which variable goes in or out of the basis, then usually no problem is encountered.

It is not known whether the critical line algorithm will cycle if two or more η_i and/or X_i go to zero simultaneously and the critical line algorithm chooses one to enter or leave the IN set by an arbitrary rule. To my knowledge, in practice such cycling has never happened. However, in case it ever does, one can build a version of the critical line algorithm that is guaranteed not to cycle, by, in effect, adding suitable powers of ϵ to the b_i and μ_i. The sequence of IN sets is the same for all sufficiently small positive ϵ. As in linear programming, it is not necessary to actually add these powers of ϵ to figure out the sequence of IN sets. A solution to the original problem can be determined from the IN sets of the perturbed problem.

A problem with unbounded E can be reduced to one with bounded E by adding the constraint

$$\mu' X \leqslant E_0. \tag{9}$$

The sequence of IN sets is the same for all sufficiently large E_0. It is not necessary to actually add constraint (9) to determine this sequence of IN sets. From them, the solution to the original problem can be inferred. The solution includes one unbounded piece (line segment in portfolio space, parabolic segment in EV space) which is efficient for

$$\lambda_E \in [\lambda^1_{\mathrm{LOW}}, \infty).$$

This completes our outline of the solution to the general portfolio problem for all possible inputs.

References

Dantzig, G. B. 1963 *Linear programming and extensions*. Princeton University Press.

Markowitz, H. M. 1956 *The optimization of a quadratic function subject to linear constraints. Nav. Res. Logistics Q.* vol. III.

Markowitz, H. M. 1987 *Mean-variance analysis in portfolio choice and capital markets*. Oxford: Basil Blackwell.

Discussion

R. Lacey (*Derivative Investment Advisers Ltd, U.K.*). How far can transaction costs analysis, described by Professor M. H. A. Davis and Dr P. Wilmott (this Volume), be incorporated into the portfolio selection method optimization module?

H. M. Markowitz. Transaction costs can be incorporated exactly if they are proportional to change in position; see Markowitz (1987). For an approximate solution when costs are linear but not proportional, see Perold (1984).

J. Plymen (*Ruislip, U.K.*). Consider the investment scene in the early 1960s when the Markowitz principles were developed. Investment statistics were rudimentary, with long term share indices confined to prices without any dividend record. Computers were too slow and expensive for any elaborate analysis. Mean-variance analysis with its crude one factor input was the only scientific technique available. Although investment inputs were developed using fundamental multifactor betas, this was only a small improvement.

United Kingdom actuaries adopted a different approach. In 1962 they developed the *Financial Times* Actuaries 500 share index. Next they set up equity market models that compare individual share performance with that of the index, obtaining a relative price ranking. (Models by Weaver & Hall, Hempsted and Clarkson have been published in actuarial journals.) With this pricing ability, portfolios are monitored at regular intervals, selling dear shares for cheap ones. This continual programme reduces risk and improves performance.

Mean-variance analysis based on more sophisticated models for performance and using semi-variance rather than mean-variance may have practical value. Mathematicians interested in finance could concentrate on: (i) actuarial market models; (ii) analytical techniques for various forms of derivatives; (iii) mathematical use of semi-variance rather than variance.

M. A. H. Dempster (*University of Essex, U.K.*). The type of index tracking portfolio management policy advocated by Mr Plymen can lead to inefficiencies with respect to any attitude-to-risk criterion, including mean-variance. As pointed out by Hodges (this Volume) the appropriate criterion for portfolio management depends on whose preferences – fund managers or ultimate beneficiaries – it embodies. In any event Professor Markowitz's recent practical portfolio experience with sophisticated mean-semi-variance methods is impressive.

R. G. Tompkins (*Kleinwort Benson Investment Management, U.K.*). In applying a mean-variance framework to emerging markets we find ridiculous results. That is, these markets have extraordinarily high historical returns and extremely low risk. This is counter-intuitive as we know how risky these markets are. We have found data series distributions to be extremely skewed and lepto Kurtic. It would be useful to expand the mean-variance framework in portfolio management to include the third and fourth moments. Perhaps this approach could be applied to the inclusion of contingent claims (such as options) into estimating optional portfolios.

Additional references

Clarkson, R. S. & Plymen, J. 1988 Improving the performance of equity portfolios. *J. Inst. Actuaries* **115**.

Perold, A. 1984 Large-scale portfolio optimization. *Man. Sci.* **30**, 1143–1160.