New insights on the mean-variance portfolio selection from de Finetti's suggestions

Flavio Pressacco and Paolo Serafini, Università di Udine

Abstract: In this paper we offer an alternative approach to the standard portfolio selection problem following the mean-variance approach to financial decisions under uncertainty originally proposed by de Finetti. Beside being a simple and natural way, alternative to the classical one, to derive the critical line algorithm, our approach throws new light on the efficient frontier and a clear characterization of its properties. We also address the problem with additional threshold constraints.

1. Introduction

It has been recently recognized (Rubinstein (2006)) that Bruno de Finetti (1940) was the first to apply the mean-variance approach in the theory of modern finance in order to solve a proportional reinsurance problem. In de Finetti's approach, as early signaled by Pressacco (1986), the reinsurance problem has much in common with the problem of asset portfolio selection, which some years later was treated in a mean-variance setting by H. Markowitz (1952, 1956).

Indeed both problems aim at minimizing the same quadratic function subject to slightly different sets of linear constraints. To solve the portfolio selection problem, but also to face the reinsurance problem in his recent critical review of de Finetti's paper, Markowitz (2006) makes recourse to the technical tools suggested by the Karush-Kuhn-Tucker optimality conditions (Karush (1939), Kuhn and Tucker (1951)), from which he derives the so called critical line algorithm.

In his paper de Finetti uses a different approach based on intuitive yet powerful ideas. In more detail he looks at the optimal set as a path in the n dimensional set of feasible retentions. The path starts from the point of largest expectation and goes on along a direction aiming to realize the largest advantage, properly measured by the ratio decrease of variance over decrease of expectation.

In order to pass from this vision to an operational procedure, de Finetti introduces a set of key functions designed to capture the benefits obtained from small feasible basic movements of the retentions. A basic movement is simply a movement of a single retention.

We showed elsewhere (Pressacco and Serafini (2007), shortly denoted as PS07) that a procedure coherently based on the key functions is able to generate the critical line algorithm and hence the whole optimal mean-variance set of a reinsurance problem.

De Finetti did not treat at all the asset portfolio selection problem. Then it is natural to raise the question whether his approach may work as an alternative way also to solve the portfolio selection problem. As a matter of strategy, the idea of looking at the efficient set as a path connecting the point of largest expectation to the point of smallest variance still maintains its validity. However, in the new portfolio setting small movements of single assets cannot be accepted as feasible basic movements. Hence we make recourse to a new type of feasible basic movements characterized by a small bilateral feasible trading between pairs of assets.

Accordingly, we define a new set of key functions designed to capture the consequences (still measured by the ratio decrease of variance over decrease of expectation) coming from such basic movements. On the basis of this set of key functions, we are able to build a procedure for the portfolio selection problem which turns out to be the exact counterpart of the one ruling the search for the mean-variance efficient set in the reinsurance case. Beside being the natural way to generate the critical line algorithm, this approach offers also a surprisingly simple and meaningful characterization of the efficient set in terms of the key functions.

Precisely, a feasible portfolio is efficient if and only if there is a non-negative value of an efficiency index such that any feasible basic bilateral trading involving two active assets has the same efficiency, while any basic bilateral trading involving one active and one non-active asset has a lower efficiency. It is interesting to note that the efficiency index of de Finetti's approach is nothing but the Lagrange multiplier of the expectation constraint in the constrained optimization setting.

In our opinion these results, that we have obtained as a proper and coherent extension of de Finetti's approach, throw new light on the intrinsic meaning and the core characteristics of the mean-variance optimum set in classical portfolio selection problems.

The same approach works also in case there are additional upper and/or lower threshold constraints, through a modification of the key functions defined now on an extended space in order to capture beside the direct also the indirect benefit coming from active constraints.

The plan of the paper is as follows. In Section 2 we briefly recall the essentials of de Finetti's approach to the reinsurance problem. In Section 3 we discuss the logic behind the introduction of the new key functions suited for the portfolio selection problem. Section 4 is devoted to the application of the new key functions approach to the portfolio selection problem in the unrealistic but didascalic case of no correlation. This opens the way to Section 5 where the application of the new approach to the general case with non-null correlation is discussed. In Section 6 we introduce the threshold constraints. Connections of our approach with classical mathematical programming is discussed in Section 7 for the standard case and in Section 8 for the case with additional constraints. Two examples follow in Section 9 before the conclusions in Section 10. An appendix is devoted to prove some formulas.

2. A recall of de Finetti's approach to the reinsurance case

We briefly recall the essentials of de Finetti's approach to the reinsurance problem. An insurance company is faced with n risks (policies). The net profit of these risks is represented by a vector of random variables with expected value $m := \{m_i > 0 : i = 1, \ldots, n\}$ and a non-singular covariance matrix $\mathbf{V} := \{\sigma_{ij} : i, j = 1, \ldots, n\}$. The company has to choose a proportional reinsurance or retention strategy specified by a retention vector x. The retention strategy is feasible if $0 \le x_i \le 1$ for all i. A retention x induces a random profit with expected value $E = x^\top m$ and variance $V = x^\top \mathbf{V}\, x$. A retention x is by definition mean-variance efficient or Pareto optimal if for no feasible retention y we have both $x^\top m \le y^\top m$ and $x^\top \mathbf{V}\, x \ge y^\top \mathbf{V}\, y$, with at least one inequality strict. Let X^* be the set of optimal retentions.

The core of de Finetti's approach is represented by the following simple and clever ideas. The set of feasible retentions is represented by points of the n dimensional unit cube. The set X^* is a path in this cube. It connects the natural starting point, the vertex $\mathbf{1}$ of full retention (with the largest expectation $E = \sum_i m_i$), to the opposite vertex $\mathbf{0}$ of full reinsurance (zero retention and hence minimum null variance). De Finetti argues that the optimum path must be the one which at any point x^* of X^* moves in such a way to get locally the largest benefit measured by the ratio decrease of variance over decrease of expectation. To translate this idea in an operational setting de Finetti introduces the so called key functions

$$F_i(x) := \frac{1}{2} \frac{\frac{\partial V}{\partial x_i}}{\frac{\partial E}{\partial x_i}} = \sum_{j=1}^{n} \frac{\sigma_{ij}}{m_i} x_j , \qquad i = 1, \ldots, n$$

Intuitively these functions capture the benefit coming at x from a small (additional or initial) reinsurance of the i-th risk. The connection between the efficient path and the key function is then straightforward: at any point of X^* one should move in such a way to provide additional or initial reinsurance only to the set of those risks giving the largest benefit (that is with the largest value of their key function). If this set is a singleton the direction of the optimum path is obvious; otherwise x should be left in the direction preserving the equality of the key functions among all the best performers. Given the form of the key functions, it is easily seen that this implies a movement on a segment of the cube characterized by the set of equations $F_i(x) = \lambda$ for all the current best performers. Here λ plays the role of the benefit parameter. And we continue on this segment until the key function of another non-efficient risk matches the current decreasing value of the benefit parameter, thus becoming a member of the best performers set. Accordingly, at this point the direction of the efficient path is changed as it is defined by a new set of equations $F_i(x) = \lambda$, with the addition of the equation for the newcomer.

A repeated sequential application of this matching logic defines the whole efficient set. De Finetti (1940) offers closed form formulas in case of no correlation and gives a largely informal sketch of the sequential procedure in case of correlated risks. As pointed out by Markowitz (2006) in his recent critical review, de Finetti overlooked the (say non-regular) case in which at some step it is not possible to find a matching point along an optimum segment before one of the currently active variables reaches a boundary value (0 or 1). We showed elsewhere (PS07) that a natural adjustment of the key functions procedure offers, also in the non-regular case, a straigthforward approach to solve the problem. Precisely, if a boundary event happens, the direction of the optimum is changed simply by forcing the boundary risk to leave the set of actively reinsured risks and freezing it at the level of full reinsurance (in case of boundary at level 0) or, less likely, at the level of full retention (in case of boundary at level 1).

It is instructive to look carefully at what happens to the key function of a risk having reached the boundary level 0. In this case of course no additional reinsurance may take place and the meaning of the key function $F_i(x)$, if $x_i = 0$, should be seen as the burden (measured by increase of variance over increase of expectation) obtained from coming back from full reinsurance. Even if this makes no sense in isolation, it should be kept in consideration in a mixture of reinsurance moves involving more than one risk. Intuitively, in order to stay frozen at level 0 and not being involved in reinsurance movements, the risk ought to be inefficient. In other words its key function value should remain at least for a while at a level of burden larger than the current level of the efficiency parameter λ.

Moreover, this makes clear that the proper economic characterization of a mean-variance efficient reinsurance choice is resumed by the following motto: "internal matching coupled with boundary dominance". Precisely, a point x is optimal if and only if there is a non-negative value of an efficiency parameter λ such that for internal risks ($0 < x_i < 1$) the key functions of all internal variables match the value of the efficiency parameter, while boundary variables are dominated, that is exhibit less efficiency. More precisely, this means $F_i(x) \leq \lambda$ for risks for which $F_i(x)$ represents a benefit, (lower benefit lower efficiency for $x_i = 1$), while $F_i \geq \lambda$ if $F_i(x)$ captures a burden (higher burden lower efficiency for $x_i = 0$). The dominance is driven by strict inequalities for any boundary variable in the internal points of the piecewise linear efficiency path, while (under a proper mild non-degeneracy condition) at the corner points of the path just one boundary variable satisfies the strict equality $F_i(x) = \lambda$.

Indeed at such corner points x, such a risk either has just matched from below (from $F_i(x) < \lambda$) or from above (from $F_i(x) > \lambda$) the set of the other previously efficient risks and is just becoming (at the corner) an efficient variable, or rather it has just reached a boundary value from an internal one and thus it is just leaving at the corner point the efficiency set becoming henceforth strongly dominated or inefficient. The distinction between those two possibilities needs a dynamic analysis of the efficient set and not merely the observation of conditions prevailing at the corner point.

An exception to this simple rule is to be found for vertex points of the unit cube where all variables are boundary variables (or the set of internal variables is empty). Then in some sense they are all inefficient, because they keep a bigger burden (if at level 0) or a lower benefit (at level 1) in comparison with (an interval of) values of the efficiency parameter λ. For further details on this point see Section 5 of PS07.

3. Key functions in the asset portfolio problem

Let us recall the essentials of the standard asset portfolio problem we first investigate in this paper. An investor is faced with n assets. The net rate of return of these assets is represented by a vector of random variables with expected value $m := \{m_i : i = 1, \ldots, n\}$ and a non-singular covariance matrix $\mathbf{V} := \{\sigma_{ij} : i, j = 1, \ldots, n\}$. The investor has a budget to invest in the given assets. It is convenient to normalize the budget to 1. Let x_i be the fraction of budget invested in the asset i. If short positions are not allowed, the portfolio strategy is feasible if $x_i \geq 0$ for all i and $\sum_i x_i = 1$. A portfolio x induces a random rate of return with expected value $E = x^\top m$ and variance $V = x^\top \mathbf{V}\, x$. A portfolio x is by definition mean-variance efficient or Pareto optimal if for no feasible portfolio y we have both $x^\top m \leq y^\top m$ and $x^\top \mathbf{V}\, x \geq y^\top \mathbf{V}\, y$, with at least one inequality strict. Let X^* be the set of optimal portfolios.

De Finetti did not treat at all (neither in his 1940 paper nor even later) the asset portfolio problem. This was later analyzed and solved by Markowitz by making recourse to technical tools provided by constrained optimization techniques meanwhile developed mainly by Kuhn and Tucker (1951).

We think that de Finetti's idea of moving along the feasible set, starting from the point of largest expectation and following the path granting at any point the largest efficiency until an end point of minimum variance is reached, is valid also in the asset portfolio problem.

However, we cannot use the same set of key functions exploited in the reinsurance problem. Indeed, beyond individual constraints, the collective constraint $\sum_i x_i = 1$ implies that benefits driven by movements of single assets, as captured by the key functions $F_i(x)$ of the reinsurance case, do not make any sense. At first glance this seems to exclude the possibility of applying de Finetti's key function approach to the asset portfolio selection problem.

On the contrary, we will show that, through a proper reformulation of key functions, it is possible to build a procedure mimicking the one suggested by de Finetti for the reinsurance case and to obtain in a natural and straightforward way something analogous to the critical line algorithm to compute the mean-variance efficient set. Furthermore, from this idea we get a simple and meaningful characterization of the efficient set.

Henceforth we will assume a labeling of the assets coherent with a strict ordering of expectations, namely $m_1 > m_2 > \ldots > m_n$. In our opinion this is not a restrictive hypothesis; after all the event of finding two assets with precisely the same expectation could be considered as one of null probability.

Recall that we want to capture through the key functions the benefit (or burden) coming from a small portfolio movement on the feasible set. In order to reach this goal let us consider for the moment simple "basic" feasible movements, that is portfolio adjustments coming from small tradings between asset i (decreasing) and asset j (increasing). More formally and with reference to a portfolio x with positive quotas of assets i and j, let us call bilateral trading in the i-j direction an adjustment of the portfolio obtained through a "small" exchange between i decreasing and j increasing. If the benefit (burden) measure is given by the ratio decrease (increase) of variance over decrease (increase) of expectation the key functions ought to be defined as:

$$F_{ij}(x) := \frac{1}{2} \frac{\frac{\partial V}{\partial x_i} - \frac{\partial V}{\partial x_j}}{\frac{\partial E}{\partial x_i} - \frac{\partial E}{\partial x_j}} = \sum_{h=1}^{n} \frac{\sigma_{ih} - \sigma_{jh}}{m_i - m_j} \, x_h$$

Note that the sign of the denominator is the same as $(j - i)$; and also that, as both i and j are active assets (that is with positive quotas) in the current portfolio x, a feasible bilateral trading may happen also in the j-i direction. Then both $F_{ij}(x)$ and $F_{ji}(x)$ describe the results of a feasible bilateral trading at x. Moreover it is immediate to check that $F_{ij}(x) = F_{ji}(x)$. Yet it is convenient to think that the economic meaning of the two functions is symmetric: precisely, if without loss of generality i is less than j, $F_{ij}(x)$ describes a benefit in algebraic sense, while $F_{ji}(x)$ describes a burden.

If in the current portfolio x_i is positive and x_j is null, then the only feasible bilateral trading may be in the direction i-j. And $F_{ij}(x)$ describes a benefit if i is less than j or a burden in the opposite case. Obviously, if both x_i and x_j are at level 0 no feasible trade between i and j may take place.

4. The standard case with null correlation

Armed with these definitions of key functions and their economic meaning, let us look at the mean-variance efficient path starting from the case of n risky asset with no correlation. This case is clearly unrealistic in an asset market, but it is nevertheless didascalic. Then the key functions simplify to

$$F_{ij}(x) = \frac{\sigma_{ii} \, x_i - \sigma_{jj} \, x_j}{m_i - m_j}$$

The starting point of the mean-variance path is $(1, 0, 0, \ldots, 0)$, the point with largest expectation (recall the ordering convention). Let us denote this starting point as $\hat{x}^2$. The choice of the index 2 may sound strange, but it is convenient because it indicates that at $\hat{x}^2$ the second asset starts to be active (see below). We leave $\hat{x}^2$ in the direction granting the largest benefit that is the largest value over j of

$$F_{1j}(\hat{x}^2) = \frac{\sigma_{11}}{m_1 - m_j}$$

(note that $\sigma_{ii} > 0$ by hypothesis). It is trivial to check that it is $j = 2$. Then the initial benefit is

$$\hat{\lambda}^2 := F_{12}(\hat{x}^2) = \frac{1}{\frac{m_1 - m_2}{\sigma_{11}}}$$

This means that the bilateral trading of the type 1-2 gives the largest benefit and dictates the efficient path leaving $\hat{x}^2$ in the direction $(-\varepsilon, \varepsilon, 0, \ldots, 0)$. The bilateral trading 1-2 remains the most efficient until we find a point on the above segment where the benefit granted by this trade is matched by another bilateral trade, that is until the nearest point, let us label it as $\hat{x}^3$, where $F_{12}(\hat{x}^3) = F_{1j}(\hat{x}^3)$ for some $j > 2$. It is easy to check that $j = 3$ and that

$$\hat{x}^3 = \frac{1}{\frac{m_1 - m_3}{\sigma_{11}} + \frac{m_2 - m_3}{\sigma_{22}}} \left(\frac{m_1 - m_3}{\sigma_{11}} \,,\, \frac{m_2 - m_3}{\sigma_{22}} \,,\, 0 \,,\, \ldots \,,\, 0 \right)$$

where the benefit is

$$\hat{\lambda}^3 := \; F_{12}(\hat{x}^3) = F_{13}(\hat{x}^3) = \frac{1}{\frac{m_1 - m_3}{\sigma_{11}} + \frac{m_2 - m_3}{\sigma_{22}}}$$

Going back to the reinsurance terminology, $\hat{x}^3$ is nothing but a corner point of the matching type. Now the matching of the two key functions F_{12} and F_{13} signals that at $\hat{x}^3$ the bilateral trade 1-2 has the same efficiency (the same benefit) of the trade 1-3. Let us comment on this point.

Remark a): at $\hat{x}^3$ also the bilateral trade 2-3 matches the same benefit $\hat{\lambda}^3$. Indeed the following result holds also in case of non-null correlation: for any triplet i, j, h of assets and any portfolio x such that at least x_i and x_j are strictly positive $F_{ij}(x) = F_{ih}(x) = \lambda$ implies $F_{jh}(x) = \lambda$. If $i < j < h$ the key functions value describes a matching of the benefits given by any bilateral trade *i-j*, *i-h*, *j-h*. The proof is straightforward (see the Corollary in Section 6).

Remark b): small feasible changes in a portfolio composition may come as well from joint movements of more than two assets, to be seen as a multilateral trading. But of course any multilateral feasible trade may be defined as a proper combination of feasible bilateral trades. For example $\delta x_i = 0.0100$, $\delta x_j = -0.0075$, $\delta x_h = -0.0025$ comes from combining $\delta x_i = 0.0075$, $\delta x_j = -0.0075$ and $\delta x_i = 0.0025$, $\delta x_h = -0.0025$; and the multilateral benefit is the algebraic combination of the bilateral benefits and if all the bilateral trades share the same benefit index then the benefit of the multilateral trade matches any combination of benefits from the bilateral trades. That is why we may concentrate on the benefits from bilateral trades neglecting the analysis of the consequences of multilateral ones.

Remark c): in some cases it could be advantageous to split a multilateral trade in bilateral components implying also one or more trading of the type *j-i* with $j > i$. Surely in isolation this cannot be a drift of movements along the efficient path, but as we shall see later it could add efficiency when inserted in the context of a multilateral trade.

Let us go back now to the point $\hat{x}^3$ where all bilateral trades between two of the first three assets and hence any feasible multilateral trade among them shares the same efficiency. At first sight there is here an embarassing lot of opportunities to leave $\hat{x}^3$ along directions granting the same benefit. But help is given by the second golden rule we receive in heritage from the reinsurance problem (PS07 p. 22): move along the path (indeed the segment) which preserves the equality of all implied key functions. In our case move along the direction $(\varepsilon_1, \varepsilon_2, \varepsilon_3, 0, \ldots, 0)$ (with $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 = 0$) which preserves the equality $F_{12}(x) = F_{13}(x)$ until a new matching point is found. It is easy to prove that (see the Appendix) a sequential application of this procedure defines a piecewise linear optimal path with corner points $\hat{x}^h$ where the asset labeled h joins the other assets $1, \ldots, h-1$ already active in the portfolio. At $\hat{x}^h$ we have

$$\hat{x}_i^h = \frac{\frac{m_i - m_h}{\sigma_{ii}}}{\sum_{j=1}^{h-1} \frac{m_j - m_h}{\sigma_{jj}}} \qquad i = 1, \ldots, h-1, \qquad \hat{x}_i^h = 0 \quad i \geq h$$

The common benefit of all bilateral tradings of the i-j type with $i < j$ is given at $\hat{x}^h$ by

$$\hat{\lambda}^h = \frac{1}{\sum_{j=1}^{h-1} \frac{m_j - m_h}{\sigma_{jj}}}$$

The ending point of the efficient path is the point

$$x_i = \frac{\frac{1}{\sigma_{ii}}}{\sum_{j=1}^{n} \frac{1}{\sigma_{jj}}}$$

of absolute minimum variance with benefit 0 for all key functions.

It is interesting to note that, going from $\hat{x}^h$ to $\hat{x}^{h+1}$, x_i is increasing (constant, decreasing) according to the negative (null, positive) sign of $\Phi_i^h := \sum_{j=1}^{h} (m_i - m_j)/\sigma_{jj}$. Note that, by decreasing λ, x_1 is surely decreasing overall the efficient path. Any other asset has increasing quotas at the beginning of its entry in portfolio and then has a behaviour with at most one relative maximum at some corner point (always by decreasing λ). The last asset surely and maybe other assets with high label may maintain definitively an increasing behaviour.

A short comment is in order regarding the beautiful simplicity of the x_i values at the corner points. For the active assets they are simply given by (a normalized to one) excess return over variance, the excess being computed with respect to the current newcomer asset.

A similar result for the no correlation case but without non-negativity constraints has been found by Lintner (Lintner 1965 p. 20-21). There the expected excess return is referred to the optimal (market) portfolio rather than to the newcomer asset.

Summing up, the case of no asset correlation shares with the twin reinsurance case the advantage of simplicity. There is an a priori labeling of the variables (but based on a different index that the one ruling the insurance case) which dictates a sequence of entry in reinsurance or respectively in portfolio. These entries happen at a sequence of matching points when a key function representing an outside variable (not yet in reinsurance or not yet in portfolio) joins the value of the key functions representing variables already in reinsurance or in portfolio. At the corner points and hence in all other points closed form formulas for the optimal portfolios (asset quotas or reinsurance quotas) and for the respective benefit are easily obtained.

As for the asset portfolio problem the necessary and sufficient condition for an internal point of the efficient path (internal to some segment) to be efficient in the no correlation case is that any bilateral trading of the i-j type ($i < j$) shares for all active assets the same benefit which is greater than the benefit coming from trading between each active i and each non-active j (here surely $i < j$). At the corner points things change only with respect to the newcomer h asset. At the corner point despite being still at level 0 it begins to share with respect to any active asset i the same benefit in bilateral trading of the i-h type, which characterize i-j trading for any pair of active assets.

5. The standard case with non-null correlation

We consider now the more realistic and important case with non-null correlation. As in the reinsurance problem, it is still convenient to distinguish between a "regular" and a non-regular case. In case of regularity corner points are always matching points and each matching point corresponds to a matching of the key function $F_{1h}(x)$ of a newcomer asset h with the common value of the key functions $F_{1i}(x)$ of the current active assets i. The only difference with the no correlation case is that the labeling induced by the expectation vector is no longer coincident with the entrance (or matching) ordering, except for the first asset of largest expectation which is still associated with the starting point. It follows that a truly sequential procedure is needed to find the whole efficient set (and a fortiori we do not have a priori closed form formulas). Of course except for the no correlation case we cannot a priori say (but only ex post) if there is regularity or not.

It is interesting to note that in any case the computational burden is quite lower than what may appear at first glance. Indeed, there is no need to compute values of the key functions $F_{ij}(x)$ for all pairs i-j (with at least i active), but it is enough (under regularity) to evaluate, for $j = 2, \ldots, n$, the $n-1$ values $F_{1j}(x)$. And the optimality condition to be checked is that there is a non negative λ such that for all active assets i we have $F_{1i}(x) = \lambda$ (note that under regularity asset 1 is surely active along the whole optimum path), while for non active ones j we have $F_{1j}(x) \leq \lambda$, with strict equality holding only at corner points just for the newcomer matching asset. This set of conditions on bilateral trading simply says that any potential bilateral trading of the 1-i type has the same non-negative benefit, which is larger than the benefit of any 1-j trading. Note that this way the asset 1 plays overall the efficient path the special role of reference asset, even if the reference asset could be more generally any active asset.

A comment is in order concerning the other feasible values $F_{ij}(x)$ of the key functions and their superfluous role in checking the optimality conditions. If both assets are active this comes from remark a) of Section 4, which grants that all bilateral tradings between two active assets share the same efficiency level λ. As for bilateral tradings between $i \neq 1$ active and j non-active, the following result holds:

$$F_{1i}(x) = \lambda \qquad \text{and} \qquad F_{1j}(x) < \lambda \qquad \text{imply} \qquad (F_{ij}(x) - \lambda)\,(j - i) < 0$$

as will be shown in the Corollary of Section 6.

This helps us to understand that the optimality conditions at x require simply that all basic bilateral tradings between each pair of active assets share the same benefit-efficiency level λ while all basic tradings between an active i and a non-active j asset have a lower efficiency level. Less efficiency means here lower benefit ($i < j$) or greater burden ($i > j$). At matching corners the matching variable becomes efficient and even if it is at the moment still non active shares the efficiency of all previously active assets.

Let us now pass to treat the non-regular case. The non-regularity comes from the fact that there is a failure in the matching sequence in the sense that along a segment of the optimum path one of the active assets reaches its boundary value 0 before a matching occurs. This is the counterpart of a boundary event in the reinsurance case and it should be clear that at a boundary point the boundary variable leaves the efficient set and remains frozen, maybe temporarily, at the level 0.

The new set of the other active assets (given that there are at least 2), determines the new direction, either preserving the equality between the key functions in case of 3 or more active assets, or in the unique feasible direction if there are only two assets. Before discussing what happens in case of only one surviving active asset we underline that the behaviour at a boundary corner is the only difference between the non regular and the regular case. As regards the computational burden it is still enough to compute at any point of the optimum path only the values of $n-1$ key functions. Yet an additional effort may be required at those boundary points, where the boundary asset played previously the role of the reference asset. Indeed its exit from the set of active assets require the choice of a new reference (the reference must be active) and then the need to compute the values of a new set of $n-1$ key functions. Hereafter the usual conditions of efficiency still hold for the new set of key functions. Let us finally discuss what happens at a boundary point where the set of active assets is a singleton, so that the point is a vertex of the n dimensional simplex of feasible allocations. This being the case, a resetting of the procedure is in order. Denoting by k the original label of the singleton asset, we look for leaving the vertex in such a way as to maximize the efficiency of a bilateral trade of the k-j type. This means looking for the largest positive value of $F_{kj}(x)$ over all $j > k$. The corresponding value of the benefit parameter could be seen as a matching value. We may say that such vertex corners are amphibious corners, that is points where both a boundary (at first) and a matching event (later) happen.

At the cost of some repetition let us summarize the essence of our alternative approach to portfolio selection, mimicking points a) to f) of p. 29-30 of PS07.

a) Given the ordering of assets coherent with expectations the procedure starts at the point (1,0,0,....) of largest expectation.

b) We leave the starting point in the direction giving the largest benefit, that is trading asset 1 with the asset j having the biggest value of $F_{1j}(x)$ at the starting point.

c) The optimum path is a piecewise linear path, whose corner points play a key role. They may be generated by two types of events: matching events and boundary events.

d) In a matching event a key function $F_{ih}(x)$, representing the equivalence class of a set of key functions capturing the burden or benefits from trading between an asset h previously non-active (with quota zero in the efficient portfolios) and one of the active (with positive quotas) assets, matches the current efficiency (benefit or burden) value given from trading between any pair of already active assets. The asset h is named matching asset. At the matching point it becomes an active asset joining the set of previous active assets. The other non-active assets j do not change their position at level zero and the values of the key functions $F_{ij}(x)$ in which they are involved reveal a lower efficiency, that is $(F_{ij}(x)-\lambda)(i-j) > 0$. We leave a matching corner point along the direction preserving the equality of the key functions concerning any internal trading that is trading within the set of active assets.

e) In a boundary event one of the previously active assets reaches the boundary value 0, before a matching event happens. When a boundary event occurs, the boundary asset leaves the set of active assets. Its level becomes fixed (maybe not definitively) at the level 0. The other risk do not change their status (either active or non-active). If there are at least three remaining active assets, we leave the corner point along the direction preserving the equality of the key functions concerning any internal trading. Note that if and only if the boundary asset is the reference asset we need to change the reference and this implies changing the set of key functions entering the computation.

If there are two active assets the direction of the internal trade is dictated by the need to decrease the expectation. If the set of active assets is the singleton i, then the corner point is a vertex of the n dimensional simplex and the one-to-one correspondence between points x^* of the mean-variance efficient set and values of the efficiency parameter λ is lost. Then in order to leave the vertex we need to decrease λ to the smallest value of the interval associated to the vertex, where one of the non-active assets j for which $j - i > 0$ realizes a new matching event, that is $F_{ij}(x) = \lambda$.

f) The procedure ends at the final corner point where any internal trading gives a null benefit $F_{ij}(x) = 0$ for any pair i, j of active assets and of course negative efficiency for any i-j trading between an active asset i and a non-active asset j.

6. A non-standard case with additional constraints

We wish to underline that some of the nice results derived up to now for the standard case rely on the assumption that there are only $n + 1$ classical constraints. Precisely, n individual non-negative constraints and one collective saturation constraint.

What about the case with additional non-classical constraints? There is a vast literature concerning the treatment through advanced optimization methods of this problem (for a good survey see Mitra et al. 2003). Here we will discuss the extension of the key function approach to a specific type of additional constraints of practical interest: upper and lower bounds on the share of portfolio globally devoted to subsets of the assets. Indeed such constraints may be often found in the statutory rules governing the allocation strategies of many financial institutions.

Maybe surprisingly, the key function approach reveals to be fully powerful in treating also the portfolio selection problem with additional upper and lower bound collective constraints.

Here is an informal description of how it works; a mathematical treatment may be found in Section 8; more detailed comments are offered in Section 9 in the discussion of Example 2. The key intuition lies in passing from an n dimensional real world to a new extended world through the addition of m shadow or virtual dimensions, one for each additional constraint. New dimensions are associated with virtual assets, whose quotas in the extended portfolio, denoted by y_p, must obey only individual non-negativity constraints. It turns out that the variables y_p are the Lagrange multipliers of the new constraints, so that if x is such that constraint p is binding, y_p is non-negative, otherwise y_p is null.

For any feasible point of the extended space $z = (x, y)$ a new set of extended key functions $G_{ij}(z)$ is defined. These functions measure the extended benefit coming from a bilateral trade of the i-j type between any pair of real world assets. In more detail, the extended benefit is the sum of the direct $F_{ij}(x)$ benefit and the indirect $H_{ij}(y)$ benefit. Here is how $H_{ij}(y)$ works: if only i is playing a role in an active constraint p, a bilateral trading of the i-j type desaturates the binding constraint and so adds an indirect benefit measured by $H_{ij}(y)$ to the direct benefit $F_{ij}(x)$. Conversely, if only j is active, the bilateral trading i-j oversaturates the already binding constraint, thus generating an indirect burden.

After that an optimum path in the extended space z is defined coherently with the application to the set of extended key functions of the rules defined for the classical case. Corner points in the extended space are generated either by

a) matching events concerning the G functions, which imply another x asset becoming active in portfolio, or

b) boundary events concerning real world assets (then leaving the portfolio) as well as virtual assets (stopping to be active), that reach the lower bound zero from positive values, or

c) a new constraint becoming binding, in which case the corresponding asset y_p begins to be positive or enters the extended portfolio.

Along the optimum path the extended benefit becomes smaller and smaller until it reaches the minimum value 0. After having defined the optimum path in the extended space we go back to the real world space, exploiting the fact that points of the optimal real world path x^* are generated by points z^* by simply projecting the optimum path from the space Z to the space X.

7. A mathematical programming formulation for the standard case

In this section we provide a mathematical foundation of the approach through key functions illustrated in the previous sections. The problem we investigate can be restated as the following quadratic problem

$$\begin{aligned} \min \quad & \frac{1}{2}\, x^T\, \mathbf{V}\, x \\ & m^T\, x \geq E \\ & \mathbf{1}^T\, x = 1 \\ & x \geq \mathbf{0} \end{aligned} \tag{1}$$

for every attainable E, i.e. $\min_i m_i \leq E \leq \max_i m_i$. The strict convexity of the objective function guarantees that there is a one-to-one correspondence between points in X^* and optimal solutions of (1) for all attainable E such that the constraint $m^T\, x \geq E$ is active. The Karush-Kuhn-Tucker conditions are necessary and sufficient for optimality of (1), since the constraints are regular and the objective function is strictly convex (see Shapiro (1979), Karush (1939), Kuhn and Tucker (1951)). The conditions are expressed through the Lagrangean function

$$L(x, \lambda, \mu, v) = \frac{1}{2}\, x^T\, \mathbf{V}\, x + \lambda\, (E - m^T\, x) + \mu\, (1 - \mathbf{1}^T\, x) - v\, x$$

and state that $\hat{x}$ is optimal if and only if there exist Lagrange multipliers $(\hat{\lambda}, \hat{\mu}, \hat{v})$, $\hat{\lambda} \geq 0$, $\hat{v} \geq 0$, such that:

$$\begin{aligned} &1) \quad \hat{x} \text{ minimizes } L(x, \hat{\lambda}, \hat{\mu}, \hat{v}) \\ &2) \quad \hat{x} \text{ is feasible in (1)} \\ &3) \quad \text{either } \hat{x}_j = 0 \text{ or } \hat{v}_j = 0 \text{ (or both) and} \\ &\qquad \text{either } m^T\, x = E \text{ or } \hat{\lambda} = 0 \text{ (or both)} \end{aligned} \tag{2}$$

In order to verify 1) of (2), since x is unconstrained (in the Lagrangean minimization), it is enough to compute:

$$\frac{\partial L}{\partial x} = \mathbf{V}\, x - \lambda\, m - \mu\, \mathbf{1} - v = 0 \tag{3}$$

i.e., componentwise

$$\sum_j \sigma_{hj}\, x_j - \lambda\, m_h - \mu - v_h = 0 \qquad h = 1, \ldots, n \tag{4}$$

We want to rephrase the optimality condition by showing how the optimal variables depend on λ. They depend also on μ, but we prefer to hide the dependence on μ by solving (3) first on μ. As before we assume that the indices are ordered as $m_1 > m_2 > \ldots > m_n$. Let k be any index such that $x_k > 0$. We denote this variable as the *reference variable.* We have $v_k = 0$ by complementarity and (4) for $h = k$ is

$$\sum_j \sigma_{kj}\, x_j - \lambda\, m_k - \mu = 0 \tag{5}$$

Now we subtract (4) from (5)

$$\sum_j (\sigma_{kj} - \sigma_{hj})\, x_j - \lambda\, (m_k - m_h) + v_h = 0 \qquad h \neq k$$

or equivalently

$$\frac{\sum_j (\sigma_{kj} - \sigma_{hj})\, x_j}{m_k - m_h} + \frac{v_h}{m_k - m_h} = \lambda \qquad h \neq k \tag{6}$$

Note that solving (6) is equivalent to solving (4). Indeed, once (6) is solved, μ can be easily computed from all other variables. We have defined the key functions as

$$F_{kh}(x) = \frac{\sum_j (\sigma_{kj} - \sigma_{hj})\, x_j}{m_k - m_h}, \qquad h \neq k$$

(note that $F_{kh}(x) = F_{hk}(x)$) and by using the key functions we may rephrase (6) as

$$F_{kh}(x) + \frac{v_h}{m_k - m_h} = \lambda \qquad h \neq k \tag{7}$$

Now we partition the variable indices $\{1, \ldots, n\} \setminus k$ into three sets as

$$I_k^* := \{h \neq k : x_h > 0\}\,, \quad I_0^k := \{h < k : x_h = 0\}\,, \quad I_k^0 := \{h > k : x_h = 0\}$$

For the sake of notational simplicity, we omit to denote that these subsets actually depend also on x. Moreover, let $I^* := I_k^* \cup \{k\}$ and $I^0 := I_0^k \cup I_k^0$ (respectively the sets of positive and null variables independently of the reference variable). Then, taking into account that $v_h \geq 0$, $m_k > m_h$, if $h \in I_k^0$ and $m_k < m_h$, if $h \in I_0^k$, the complementarity condition can be restated through the key functions in the following form:

Optimality condition: *Let k such that $x_k > 0$. Then $x \geq 0$ is optimal if and only if $\mathbf{1}^T x = 1$ and there exists $\lambda \geq 0$ such that*

$$F_{kh}(x) \geq \lambda, \quad h \in I_0^k, \qquad F_{kh}(x) = \lambda, \quad h \in I_k^*, \qquad F_{kh}(x) \leq \lambda, \quad h \in I_k^0 \tag{8}$$

∎

The following facts can be deduced from the optimality condition:

Corollary: *Let $i \in I_k^*$. If $j \in I_k^*$ then $F_{ij}(x) = \lambda$. If $j \in I^0$ then*

$$F_{ij}(x) \le \lambda \quad \textit{if} \;\; i < j, \qquad F_{ij}(x) \ge \lambda \quad \textit{if} \;\; i > j$$

Proof: : Let us denote $\varphi_i := \sum_j \sigma_{ij}\, x_j$. So we may rewrite

$$F_{kh}(x) = \frac{\varphi_k - \varphi_h}{m_k - m_h}$$

Hence

$$F_{ki}(x) = F_{kj}(x) = \lambda \quad \Longrightarrow \quad (\varphi_k - \varphi_i) = \lambda\,(m_k - m_i), \qquad (\varphi_k - \varphi_j) = \lambda\,(m_k - m_j)$$

from which we easily derive $(\varphi_i - \varphi_j) = \lambda\,(m_i - m_j)$, i.e., $F_{ij}(x) = \lambda$. For $i \in I_k^*$, $j \in I^0$ we have

$$F_{ki}(x) = F_{kj}(x) + \frac{v_j}{m_k - m_j} = \lambda \quad \Longrightarrow$$

$$(\varphi_k - \varphi_i) = \lambda\,(m_k - m_i), \qquad (\varphi_k - \varphi_j) + v_j = \lambda\,(m_k - m_j)$$

from which we derive

$$(\varphi_i - \varphi_j) + v_j = \lambda\,(m_i - m_j) \quad \Longrightarrow \quad \frac{\varphi_i - \varphi_j}{m_i - m_i} + \frac{v_j}{m_i - m_j} = F_{ij}(x) + \frac{v_j}{m_i - m_j} = \lambda$$

and the thesis follows. ∎

This result implies that the role of reference variable can be subsumed by any variable in I_k^* without changing the optimality conditions, provided the sets I_k^*, I_0^k and I_k^0 are duly redefined according to the new reference, and, more importantly, for the same value of λ. In other words resetting the reference does not affect the value of λ.

The set I_k^* can be empty only in the extreme cases $x_k = 1$ and $x_h = 0$, $h \neq k$. In this case (8) becomes

$$F_{kh}(x) = \frac{\sigma_{kk} - \sigma_{hk}}{m_k - m_h} \le \lambda, \quad h > k, \qquad F_{kh}(x) = \frac{\sigma_{kk} - \sigma_{hk}}{m_k - m_h} \ge \lambda, \quad h < k,$$

Hence, if

$$\max\left\{\max_{h>k} \frac{\sigma_{kk} - \sigma_{hk}}{m_k - m_h}\;;\;0\right\} \le \min_{h<k} \frac{\sigma_{kk} - \sigma_{hk}}{m_k - m_h} \tag{9}$$

the point $x_k = 1$ and $x_h = 0$, $h \neq k$, is optimal with λ taking any value within the above interval. Note that the point $(1, 0, \ldots, 0)$ is always optimal (since the r.h.s. term is missing) and that $x_k = 1$ and $x_h = 0$, $h \neq k$, can be optimal only if $\sigma_{kk} < \sigma_{hk}$ for all $h < k$ (necessary but not sufficient condition). In particular the point $(0, 0, \ldots, 1)$ of absolute minimum mean can be also mean-variance efficient if and only if $\sigma_{nn} < \sigma_{hn}$ for all h. In this case it is the end point of the set X^*.

If the set I_k^* is not empty the optimality condition $F_{kh}(x) = \lambda$, $h \in I_k^*$, is a linear system in the variables in I^*:

$$\sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h}\, x_j = \lambda, \qquad h \in I_k^* \tag{10}$$

Adding the condition $\sum_{j \in I^*} x_j = 1$ yields a square linear system whose solution is an affine function of λ

$$x_h := w_h + \lambda\, z_h, \qquad h \in I^* \tag{11}$$

(with w solution of the linear system with r.h.s. $(0, 0, \ldots, 0, 1)$ and z solution with r.h.s. $(1, 1, \ldots, 1, 0)$) and clearly $x_h = 0$, $h \in I^0$.

As stated in PS07 the minimum portfolio variance in (1) is a strictly convex monotonically increasing function of the mean E and the multiplier λ is its derivative (or a subgradient on the points of non-differentiability). Therefore the set X^* can be parametrized via λ instead of E, taking into account that some points of X^*, where the derivative of the function has a discontinuity jump, correspond to an interval of values for λ.

Basing on the key functions we sketch a computational procedure, analogous to the critical line algorithm by Markowitz, to describe X^* parametrized via λ.

We first find the range of values λ for which x as computed in (11) is optimal. According to the optimality condition, for the variables in I_k^0 we must have

$$\sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h}\, x_j \leq \lambda, \qquad h \in I_k^0$$

i.e.

$$\sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h}\, (w_j + \lambda\, z_j) \leq \lambda, \qquad h \in I_k^0$$

$$\sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h}\, w_j \leq \lambda\, (1 - \sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h}\, z_j), \qquad h \in I_k^0$$

Let

$$\beta_h := \sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h}\, w_j, \qquad \gamma_h := (1 - \sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h}\, z_j)$$

then

$$\lambda_1 := \max_{h \in I_k^0 : \gamma_h > 0} \frac{\beta_h}{\gamma_h} \leq \lambda \leq \min_{h \in I_k^0 : \gamma_h < 0} \frac{\beta_h}{\gamma_h} \tag{12}$$

Similarly for the variables in I_0^k we must have

$$\sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h}\, x_j \geq \lambda, \qquad h \in I_0^k$$

i.e.

$$\lambda_2 := \max_{h \in I_0^k : \gamma_h < 0} \frac{\beta_h}{\gamma_h} \leq \lambda \leq \min_{h \in I_0^k : \gamma_h > 0} \frac{\beta_h}{\gamma_h} \tag{13}$$

Furthermore we must have

$$x_h = w_h + \lambda\, z_h \geq 0, \qquad h \in I^* \quad \Longrightarrow \quad \lambda_3 := \max_{h \in I^* : z_h > 0} -\frac{w_h}{z_h} \leq \lambda \leq \min_{h \in I^* : z_h < 0} -\frac{w_h}{z_h} \tag{14}$$

As in PS07 the computational procedure starts from the optimal point $x^2 := (1, 0, 0, \ldots, 0)$ and reference $k = 1$. Then we define the first corner value for λ as (according to (9))

$$\lambda^2 := \max \left\{ \max_{h > 1} \frac{\sigma_{11} - \sigma_{h1}}{m_1 - m_h} \; ; \; 0 \right\} \tag{15}$$

If $\lambda^2 = 0$ the procedure stops and X^* consists of the single point $(1, 0, 0, \ldots, 0)$. Otherwise let h be an index for which the maximum in (15) is attained. Then we define $I_k^* := \{h\}$, $I_0^k := \emptyset$ and $I_k^0 := I \setminus I^*$. Then we recursively carry out the following computations by decreasing λ from $\lambda = \lambda^2$ until $\lambda = 0$. Let r be an index identifying the corner values. We initially set $r := 3$.

According to the previous results, for a given reference k and given sets I_k^*, I_0^k, I_k^0, we first compute the vectors w and z by solving a linear system, then we compute the values λ_1, λ_2 and λ_3. As we have seen λ can be decreased up to the corner value

$$\lambda^r := \max \{\lambda_1 \; ; \; \lambda_2 \; ; \; \lambda_3 \; ; \; 0\}$$

associated to the corner optimal point (as from (11))

$$\hat{x}_h^r := w_h + \lambda^r \, z_h, \qquad h \in I^*, \qquad\qquad \hat{x}_h^r := 0, \qquad h \in I^0$$

The sets I_k^*, I_0^k, I_k^0 are updated as:

– if the maximum is λ_1, then the index yielding the maximum moves from I_0^k to I_k^*;

– if it is λ_2, then the index yielding the maximum moves from I_k^0 to I_k^*. In both cases this is a matching event and no change of reference variable is necessary;

– if it is λ_3 it is a boundary event and we have to distinguish two cases: i) the boundary variable is the reference variable and we need to reset the reference variable; ii) the boundary variable is not the reference variable and this simply implies a movement of its index from I_k^* to either I_0^k or I_k^0 (according to the index). Furthermore, if in either case the reference variable (possibly after resetting) has value 1 and consequently all other variables are zero, then λ has a discontinuity jump between the extremes of the interval specified by (9)

$$\underline{\lambda}^r := \max \left\{ \max_{h>k} \frac{\sigma_{kk} - \sigma_{hk}}{m_k - m_h} \; ; \; 0 \right\}, \qquad \overline{\lambda}^r := \min_{h<k} \frac{\sigma_{kk} - \sigma_{hk}}{m_k - m_h} = \lambda^r$$

– finally if the maximum is 0 (or $\underline{\lambda}^r = 0$) the computation stops and we have reached the end point of X^*. Otherwise r is increased by one and the computation is iterated.

Let us briefly address the question of a risk free asset possibly present in the portfolio. Such an asset (let us label it as f) is characterized by a constant rate of return m_f and covariances $\sigma_{fj} = \sigma_{jf} = 0$ for all j. If a risk free asset becomes active then in the linear system (10) the coefficient of x_f is null. This means that all other variables in I^* are a linear function of λ and not just an affine function. So they decrease as λ decreases by keeping fixed their mutual ratios. The variable x_f increases to make the sum equal to 1. This is the counterpart for the case with non-negativity constraints of results, well known in classical portfolio theory literature under the name of "two funds theorems" or "separation theorems", going back to Tobin (1958) (see also Lintner (1965)). Note that in (11) $w_h = 0$, $h \in I^*$, and so $\beta_h = 0$ and $\lambda_1 = 0$ in (12). Consequently no variables in I^0 enter I^*. When $\lambda = 0$ the computation stops with the portfolio consisting entirely of the risk free asset, clearly the point of minimum null variance on X^* and also the ending point.

Note that the optimality condition is always verified by a portfolio $\bar{x}$ consisting entirely of the risk free asset. Indeed we have

$$F_{fh}(\bar{x}) = \frac{\sigma_{ff} - \sigma_{hf}}{m_f - m_h} = 0 \leq \lambda, \quad h > f, \qquad F_{fh}(\bar{x}) = 0 \geq \lambda, \quad h < f$$

from which $\bar{x}$ is optimal with $\lambda = 0$. This means that the risk free asset becomes active at a particular value of $\bar{\lambda}$ and then for $0 \le \lambda \le \bar{\lambda}$ the portfolio of the other assets remains fixed and only the budget allocated to them decreases. The portfolio $\bar{x}$ can be also optimal with $\lambda > 0$ if there are no indices $h < f$. It is intuitively clear that if the risk free asset has the highest return, an unlikely practical case, the only optimal point is $\bar{x}$ for any value of λ.

8. A mathematical programming formulation with additional constraints

We now suppose that a family $\mathcal{U}$ of asset subsets is given together with bounds $0 < u(J) < 1$ for each $J \in \mathcal{U}$. The following saturation constraints must be satisfied by a feasible portfolio

$$\sum_{j \in J} x_j \le u(J), \quad J \in \mathcal{U} \tag{16}$$

Let $q := |\mathcal{U}|$. We denote the q constraints (16) as

$$e^{p\top} x \le c_p, \qquad p = 1, \dots, q$$

where, if p refers to the set J in the family $\mathcal{U}$, $c_p = u(J)$, $e^p_j = 1$ if $j \in J$ and 0 otherwise.

Sometimes, feasible portfolios must obey also constraints in which the sum of certain specified assets is lower bounded by a positive constant. Since all assets sum up to one, this constraint can be equivalently expressed as an upper bound on the sum of the complement set of assets. Therefore, without loss of generality, we assume that all additional constraints are of the form (16).

Then the portoflio selection problem can be stated as:

$$\begin{aligned} \min \quad & \frac{1}{2} x^\top \mathbf{V} x \\ & m^\top x \ge E \\ & \mathbf{1}^\top x = 1 \\ & e^{p\top} x \le c_p \qquad p = 1, \dots, q \\ & x \ge \mathbf{0} \end{aligned}$$

By applying the KKT conditions (with y_p dual variables of the additional constraints) we get

$$\sum_j \sigma_{hj} x_j - \lambda m_h - \mu + \sum_p y_p e^p_h - v_h = 0 \qquad h = 1, \dots, n$$

and

$$\sum_j (\sigma_{kj} - \sigma_{hj}) x_j - \lambda (m_k - m_h) + \sum_p y_p (e^p_k - e^p_h) + (v_h - v_k) = 0 \qquad h \ne k$$

We define an *extended key function* as

$$G_{kh}(x, y) := \frac{\sum_j (\sigma_{kj} - \sigma_{hj}) x_j}{m_k - m_h} + \frac{\sum_p (e^p_k - e^p_h) y_p}{m_k - m_h} = F_{kh}(x) + H_{kh}(y)$$

We postpone a discussion on the meaning of the variables y_p and the functions $H_{kh}(y)$ to the end of the section.

Note that $G_{kh}(x,y) = G_{hk}(x,y)$. Let k be a reference asset such that $x_k > 0$, so that $v_k = 0$. Then we define as in the previous case

$$I_k^* := \{h \neq k : x_h > 0\}, \quad I_0^k := \{h < k : x_h = 0\}, \quad I_k^0 := \{h > k : x_h = 0\}$$

$$I^* := I_k^* \cup \{k\} \qquad I^0 := I_0^k \cup I_k^0$$

In addition we define

$$Q^* := \left\{p : e^{p\top} x = c_p\right\}, \qquad Q^0 := \left\{p : e^{p\top} x < c_p\right\}$$

and, as before, we omit to denote that Q^* and Q^0 depend on x. Q^* and Q^0 are the sets of active and non-active constraints respectively. Then from

$$G_{kh}(x,y) + \frac{v_h}{m_k - m_h} = \lambda \qquad h \neq k$$

we get the following optimality conditions, which have the same form as in the standard case.

Extended optimality conditions: *Let k such that $x_k > 0$. Then $(x,y) \geq 0$ are optimal if and only if x is feasible, $y_p = 0$, for $p \in Q^0$, and there exists $\lambda \geq 0$ such that*

$$G_{kh}(x,y) \geq \lambda, \quad h \in I_0^k, \qquad G_{kh}(x,y) = \lambda, \quad h \in I_k^*, \qquad G_{kh}(x,y) \leq \lambda, \quad h \in I_k^0$$

∎

The following corollary is valid also in the extended case.

Corollary: *Let $i \in I_k^*$. If $j \in I_k^*$ then $G_{ij}(x,y) = \lambda$. If $j \in I^0$ then*

$$G_{ij}(x,y) \leq \lambda \quad \text{if } \; i < j, \qquad G_{ij}(x,y) \geq \lambda \quad \text{if } \; i > j$$

∎

Note that $G_{kh}(x,y) = \lambda$ together with the sum one constraint and the active constraints is the following square linear system in the active x and y variables:

$$\begin{aligned} &F_{kh}(x) + H_{kh}(y) = \lambda && h \in I_k^* \\ &e^{p\top} x \qquad\qquad\quad = c_p && p \in Q^* \\ &\mathbf{1}^\top x \qquad\qquad\quad\; = 1 && \end{aligned} \tag{17}$$

The solution is affinely parametrized by λ as

$$x_h := w_h + \lambda\, z_h, \qquad h \in I^*, \qquad\quad y_p = w_p + \lambda\, z_p, \qquad p \in Q^* \tag{18}$$

(with w solution of the linear system with r.h.s. $(0, 0, \ldots, c, 1)$ and z solution with r.h.s. $(1, 1, \ldots, 0, 0)$). Clearly $x_h = 0$, $h \in I^0$, $y_p = 0$, $p \in Q^0$.

The range of values λ for which x and y as computed in (18) are optimal can be computed as in the previous case with some new conditions. According to the optimality condition, for the variables in $h \in I_k^0$

we must have $G_{kh}(x,y) \le \lambda$ and for the variables in $h \in I_0^k$ we must have $G_{kh}(x,y) \ge \lambda$. With respect to the previous case we just need redefining

$$\beta_h := \sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h} w_j + \sum_{p \in Q^*} \frac{(e_k^p - e_h^p)}{m_k - m_h} w_p, \qquad \gamma_h := (1 - \sum_{j \in I^*} \frac{\sigma_{kj} - \sigma_{hj}}{m_k - m_h} z_j - \sum_{p \in Q^*} \frac{(e_k^p - e_h^p)}{m_k - m_h} z_p)$$

Then

$$\lambda_1 := \max_{h \in I_k^0 : \gamma_h > 0} \frac{\beta_h}{\gamma_h} \le \lambda \le \min_{h \in I_k^0 : \gamma_h < 0} \frac{\beta_h}{\gamma_h} \tag{19}$$

$$\lambda_2 := \max_{h \in I_0^k : \gamma_h < 0} \frac{\beta_h}{\gamma_h} \le \lambda \le \min_{h \in I_0^k : \gamma_h > 0} \frac{\beta_h}{\gamma_h} \tag{20}$$

From $x_h = w_h + \lambda\, z_h \ge 0$, for $h \in I^*$ we have

$$\lambda_3 := \max_{h \in I^* : z_h > 0} -\frac{w_h}{z_h} \le \lambda \le \min_{h \in I^* : z_h < 0} -\frac{w_h}{z_h} \tag{21}$$

Now we also need $y_p = w_p + \lambda\, z_p \ge 0$, for $p \in Q^*$, which implies

$$\lambda_4 := \max_{p \in Q(x) : z_p > 0} -\frac{w_p}{z_p} \le \lambda \le \min_{p \in Q(x) : z_p < 0} -\frac{w_p}{z_p} \tag{22}$$

and $e^{p\top} x \le c_p$ for $p \in Q^0$, so that $e^{p\top} w + \lambda\, e^{p\top} z \le c_p$ and therefore

$$\lambda_5 := \max_{p \in Q^0 : e^{p\top} z < 0} \frac{c_p - e^{p\top} w}{e^{p\top} z} \le \lambda \le \min_{p \in Q^0 : e^{p\top} z > 0} \frac{c_p - e^{p\top} w}{e^{p\top} z}$$

In this case the computational procedure starts from the optimal point $x^2 := (1, 0, 0, \dots, 0)$ if and only if it is a feasible point. In general we compute the starting point from

$$\begin{aligned} \max \quad & m^\top x \\ & \mathbf{1}^\top x = 1 \\ & e^{p\top} x \le c_p \qquad p = 1, \dots, q \\ & x \ge \mathbf{0} \end{aligned} \tag{23}$$

Let $\hat{x}$ be the optimal solution of (23) and let k, I_k^* and Q^* be defined accordingly. Then we solve the following linear programming problem with respect to λ and the active y_p

$$\begin{aligned} \lambda^2 := \min \quad & \lambda \\ & G_{kh}(\hat{x}, y) = \lambda \qquad h \in I_k^* \\ & G_{kh}(\hat{x}, y) \le \lambda \qquad h \in I^0 \\ & y_p = 0, \qquad\qquad p \in Q^0 \\ & y_p \ge 0, \qquad\qquad p \in Q^* \end{aligned} \tag{24}$$

If $\lambda^2 = 0$ the procedure stops and X^* consists of the single point $\hat{x}$. Otherwise, if there exists and index $h \in I^0$ such that at optimality in (24) $G_{kh}(\hat{x}, y) = \lambda$, we reset $I^* := I^* \cup h$ and I_0^k, I_k^0 accordingly, and, if there exists and index $p \in Q^*$ such that at optimality in (24) $y_p = 0$, we reset $Q^* := Q^* \setminus p$, $Q^0 := Q^0 \cup p$. Then we recursively carry out the following computations by decreasing λ from $\lambda = \lambda^2$ until $\lambda = 0$. Let r be an index identifying the corner values. We initially set $r := 3$.

We first compute the vectors w and z by solving the linear system (17), then we compute the values λ_1, λ_2, λ_3, λ_4 and λ_5. As we have seen λ can be decreased up to the corner value

$$\lambda^r := \max\{\lambda_1\ ;\ \lambda_2\ ;\ \lambda_3\ ;\ \lambda_4\ ;\ \lambda_5\ ;\ 0\}$$

associated to the corner optimal point (as from (11))

$$\begin{array}{llll} \hat{x}_h^r := w_h + \lambda^r z_h, & h \in I^*, & \hat{x}_h^r := 0, & h \in I^0, \\ \hat{y}_p^r := w_p + \lambda^r z_p, & p \in Q^* & \hat{y}_p^r := 0 & p \in Q^0 \end{array}$$

The sets I_k^*, I_0^k, I_k^0, Q^*, Q^0 are updated as:

– if the maximum is λ_1, then the index yielding the maximum moves from I_0^k to I_k^*;

– if it is λ_2, then the index yielding the maximum moves from I_k^0 to I_k^*. In both cases this is a matching event (with respect to the G and not the F functions) and no change of reference variable is necessary;

– if it is λ_3 it is a boundary event, regarding a x variable, and we have to distinguish two cases: i) the boundary variable is the reference variable and we need to reset the reference variable; ii) the boundary variable is not the reference variable and this simply implies a movement of the variable from I_k^* to either I_0^k or I_k^0 (according to the index). Furthermore, if in either case the reference variable (possibly after resetting) has value 1 and consequently all other variables are zero (and we expect $Q^* = \emptyset$), then λ has a discontinuity jump between the extremes of the interval specified by (9)

$$\underline{\lambda}^r := \max\left\{\max_{h>k} \frac{\sigma_{kk} - \sigma_{hk}}{m_k - m_h}\ ;\ 0\right\}, \qquad \overline{\lambda}^r := \min_{h<k} \frac{\sigma_{kk} - \sigma_{hk}}{m_k - m_h} = \lambda^r$$

– if it is λ_4 then the index yielding the maximum moves Q^* to Q^0. This is a boundary event for a y variable and the corresponding constraint stops being active;

– if it is λ_5 then the index yielding the maximum moves Q^0 to Q^*. This is a boundary event regarding a set of x variables which saturate the constraint and its effect is like a matching event for a y variable which enters the portfolio;

– finally if the maximum is 0 (or $\underline{\lambda}^r = 0$) the computation stops and we have reached the end point of X^*. Otherwise r is increased by one and the computation is iterated.

We now provide an informal discussion on the possible meaning we may assign to the variables y_p and the functions $H_{kh}(y)$. Without additional constraints, the key functions $F_{kh}(x)$ give the efficiency level at which trades between active assets can be carried out at the best trade-off between mean and variance. Pairs k-h not at that level cannot be efficiently traded. The presence of additional constraints breaks this invariance among active assets and efficient trades can take place at different key function values that crucially depend on the constraints.

However, the invariance is recovered if we enlarge the asset set to include 'virtual' assets corresponding to the constraints. The variables y_p are the virtual asset values and the functions $H_{kh}(y)$ take care of the indirect efficiency loss or gain due to the additional constraints. This provides new efficiency levels given by $F_{kh}(x) + H_{kh}(y)$, which is the sum of the direct and the indirect benefit or burden, and it is precisely these values which are invariant among the active assets. The variable λ refers now to this composite index and, thanks to the invariance, can be used to parametrize the mean-variance efficient set as in the standard case.

Looking in detail the extended key function $G_{kh}(x,y)$ we see that, if both assets k and h are present in the constraint p or if they are not, then the coefficient $e_k^p - e_h^p$ is null and there is no need to take care of the virtual asset p.

But, if the asset k is in the constraint and h is not, then part of the efficiency is subsumed by the virtual asset value y_p. Intuitively, if only i is playing a role in p, a bilateral trade of the i-j type weakens the binding constraint which may add a virtual benefit measured by $(e_i^p - e_j^p)\, y_p/(m_i - m_j)$ to the real world benefit $F_{ij}(x)$. Conversely if only j is active the bilateral trading i-j strenghtens the already binding constraint thus generating a virtual burden. See also the Example 2 in the next Section.

9. Examples

Example 1. The first example considers a problem without additional constraints. Let

$$n = 3, \qquad m = \begin{pmatrix} 6 & 5 & 1 \end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix} 30 & 6 & -1 \\ 6 & 5 & 2 \\ -1 & 2 & 2 \end{pmatrix}$$

The key functions are:

$$F_{12}(x) = 24\,x_1 + x_2 - 3\,x_3, \qquad F_{13} = \frac{31}{5}\,x_1 + \frac{4}{5}\,x_2 - \frac{3}{5}\,x_3, \qquad F_{23} = \frac{7}{4}\,x_1 + \frac{3}{4}\,x_2$$

Starting with $x^2 = (1,0,0)$ and $k = 1$ we have

$$F_{12}(1,0,0) = 24, \qquad F_{13}(1,0,0) = \frac{31}{5}$$

so that we start with $I^* = \{1,2\}$, $I_k^0 = \{3\}$, $I_0^k = \emptyset$, $\lambda^2 = 24$. This means that the optimal path starts with the direction given by trades only between 1 and 2. From $F_{12}(x_1, x_2) = \lambda$, $x_1 + x_2 = 1$, we get

$$x_1 = \frac{\lambda - 1}{23}, \qquad x_2 = \frac{24 - \lambda}{23}$$

With these values the optimality condition $F_{13}(x) \leq \lambda$ implies $\lambda \geq 65/88$, and the non-negativity condition of the variables implies $\lambda \geq 1$. Then the next corner point is given by $\lambda^3 = 1$ where we have $x^3 = (0,1,0)$. This is a boundary point of vertex type and we need resetting the reference variable to $k = 2$. Now λ must be decreased up to $F_{23}(0,1,0) = 3/4$, where we have $I^* = \{2,3\}$, $I_k^0 = \emptyset$, $I_0^k = \{1\}$. From $F_{23}(x_2, x_3) = \lambda$ and $x_2 + x_3 = 1$ we get

$$x_2 = \frac{4\,\lambda}{3}, \qquad x_3 = \frac{3 - 4\,\lambda}{3}$$

With these values the optimality condition $F_{21}(x) \geq \lambda$ implies $\lambda \geq 9/13$ and the non-negativity condition gives $\lambda \geq 0$. Hence the next breakpoint is for $\lambda^4 = 9/13$ with matching point $x^4 = (0, 12/13, 1/13)$. At this point the variable x_1 joins the set I^* and we need to solve $F_{21}(x) = F_{23}(x) = \lambda$ together with $x_1 + x_2 + x_3 = 1$, from which we get

$$x_1 = \frac{9 - 13\,\lambda}{53}, \qquad x_2 = \frac{101\,\lambda - 21}{53}, \qquad x_3 = \frac{65 - 88\,\lambda}{53}$$

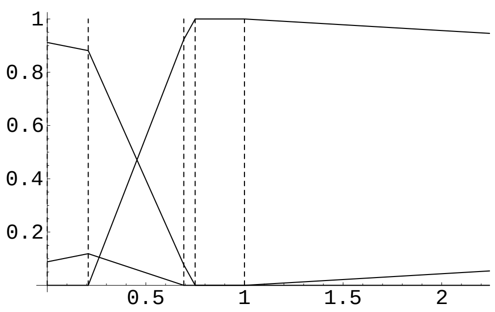
At $\lambda^5 = 21/101$ and point $x^5 = (12/101, 0, 89/101)$ the variable x_2 becomes null and we have to reset the reference to $k = 1$. Now we need to solve $F_{13}(x) = \lambda$, $x_1 + x_3 = 1$, from which we get

$$x_1 = \frac{3 + 5\,\lambda}{34}, \qquad x_3 = \frac{31 - 5\,\lambda}{34}$$

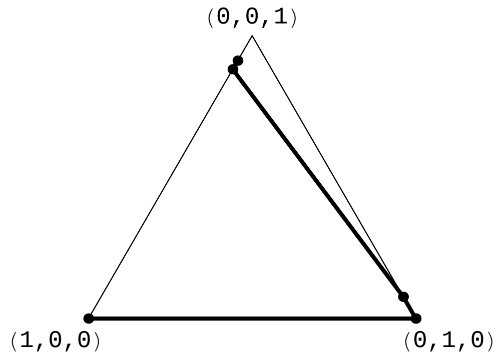
Now λ can be decreased up to 0 so that the final optimal point is

$$x^6 = \left\{ \frac{3}{34} \;;\; 0 \;;\; \frac{31}{34} \right\}$$

∎



Example 1: x_i as a function of λ (not displayed for all λ)



Example 1: X^*

Example 2. This second example shows how to deal with additional constraints. For the sake of simplicity we consider assets with no correlation:

$$m = \{10, 5, 2, 1\}, \qquad \mathbf{V} = \mathrm{diag}\,\{5, 3, 2, 1\}$$

and constraints

$$x_2 + x_3 \le \frac{1}{3}, \qquad x_3 + x_4 \le \frac{1}{3}, \qquad x_1 \le \frac{3}{4}$$

The extended key functions are:

$$\begin{aligned}
G_{12}(x,y) &= x_1 - \frac{3}{5}\,x_2 - \frac{1}{5}\,y_1 + \frac{1}{5}\,y_3 \\
G_{13}(x,y) &= \frac{5}{8}\,x_1 - \frac{1}{4}\,x_3 - \frac{1}{8}\,y_1 - \frac{1}{8}\,y_2 + \frac{1}{8}\,y_3 \\
G_{14}(x,y) &= \frac{5}{9}\,x_1 - \frac{1}{9}\,x_4 - \frac{1}{9}\,y_2 + \frac{1}{9}\,y_3 \\
G_{23}(x,y) &= x_2 - \frac{2}{3}\,x_3 - \frac{1}{3}\,y_2 \\
G_{24}(x,y) &= \frac{3}{4}\,x_2 - \frac{1}{4}\,x_4 + \frac{1}{4}\,y_1 - \frac{1}{4}\,y_2 \\
G_{34}(x,y) &= 2\,x_3 - x_4 + y_1
\end{aligned}$$

The starting point in the space X, as computed from (23), is

$$\hat{x}_1 = \frac{3}{4}, \qquad \hat{x}_2 = \frac{1}{4}, \qquad \hat{x}_3 = 0, \qquad \hat{x}_4 = 0$$

By taking $k = 1$ as the reference variable we have $I_k^* = \{2\}$. The third constraint is active, so $Q^* = \{3\}$, $\hat{y}_1 = \hat{y}_2 = 0$. We only need to consider the key functions $G_{12}(x,y)$, $G_{13}(x,y)$, $G_{14}(x,y)$, which at the starting point are:

$$G_{12}(x,y) = \frac{3}{5} + \frac{1}{5}\,y_3, \qquad G_{13}(x,y) = \frac{15}{32} + \frac{1}{8}\,y_3, \qquad G_{14}(x,y) = \frac{5}{12} + \frac{1}{9}\,y_3$$

Therefore (24) is

$$\begin{aligned} \min \quad & \lambda \\ & \lambda = \frac{3}{5} + \frac{1}{5}\, y_3 \\ & \lambda \geq \frac{15}{32} + \frac{1}{8}\, y_3 \\ & \lambda \geq \frac{5}{12} + \frac{1}{9}\, y_3 \\ & y_3 \geq 0, \end{aligned}$$

with $\hat{y}_3 = 0$ and $\hat{\lambda} = 3/5$. The starting point in the space Z is therefore $\hat{z} = (3/4, 1/4, 0, 0; 0, 0, 0)$. The fact that $\hat{y}_3 = 0$ indicates that the point is degenerate. Indeed, by solving $G_{12}(x_1, x_2, y_3) = \lambda$, $x_1 = 3/4$, $x_1 + x_2 = 1$, we get $y_3 = 0$, so that the third constraint becomes inactive already at the start. Therefore the first segment of the optimal path is computed from $G_{12}(x_1, x_2) = \lambda$ and $x_1 + x_2 = 1$ giving

$$x_1 = \frac{3}{8} + \frac{5\,\lambda}{8}, \qquad x_2 = \frac{5}{8} - \frac{5\,\lambda}{8}$$

The efficiency of the trade 1-2 (measured by λ) is due only to the direct benefit $F_{12}(x_1, x_2) = \lambda$, since $H(y) = 0$. By decreasing λ we find, at $\lambda = 7/15$ and $x_1 = 2/3$, $x_2 = 1/3$, a boundary event for the first constraint which becomes active. This is also a vertex point, so that, by solving

$$G_{12}(x_1, x_2, y_1) = \lambda, \qquad x_2 = \frac{1}{3}, \qquad x_1 + x_2 = 1$$

x_1 and x_2 remain fixed while y_1 and λ change according to $y_1 = 7/3 - 5\,\lambda$. As can be seen a decrease of λ from the value 7/15 leads to a stop in the trade 1-2. The presence of the constraint does not allow trading at the current efficiency and we have to decrease the efficiency index before real world trading can restart. The efficiency loss is measured by the function $H_{12}(y)$ in such a way to balance the loss of the direct benefit key function $F_{12}(x)$.

Note also that the point $(2/3, 1/3, 0, 0)$ is a vertex of the feasible set (in the X space). This case is similar to the vertices $(0, \ldots, 0, 1, 0, \ldots, 0)$ in the classical case, where a discontinuity jump of the variables takes place. Also in this case the real asset variables x do not change for an interval of values of λ. However, the jump is not discontinuous for the virtual assets y variables which measure smoothly the effect of the constraints on the efficiency level.

By decreasing λ we find at $\lambda = 10/27$ and $y_1 = 13/27$ a matching event (due only to the indirect contributions H) with asset 4 becoming active.

By computing $G_{12}(x_1, x_2, x_4, y_1) = G_{14}(x_1, x_2, x_4, y_1) = \lambda$, $x_2 = 1/3$, $x_1 + x_2 + x_4 = 1$ we get

$$x_1 = \frac{1}{9} + \frac{3\,\lambda}{2}, \quad x_2 = \frac{1}{3}, \quad x_4 = \frac{5}{9} - \frac{3\,\lambda}{2}, \quad y_1 = -\frac{4}{9} + \frac{5\,\lambda}{2}$$

Now a trade 1-4 is taking place. Its efficiency is only due to the direct part F_{14} since $H_{14} = 0$ so that $G_{14} = F_{14} = \lambda$.

By decreasing λ we find a matching event due to asset 3 becoming active. The breakpoint is

$$\lambda = \frac{1}{3}, \quad x_1 = \frac{11}{18}, \quad x_2 = \frac{1}{3}, \quad x_4 = \frac{1}{18}, \quad y_1 = \frac{7}{18}$$

By computing $G_{12}(x,y_1) = G_{13}(x,y_1) = G_{14}(x,y_1) = \lambda$, $x_2 + x_3 = 1/3$, $x_1 + x_2 + x_3 + x_4 = 1$ we get

$$x_1 = \frac{1}{9} + \frac{3\,\lambda}{2}, \quad x_2 = \frac{2}{15} + \frac{3\,\lambda}{5}, \quad x_3 = \frac{1}{5} - \frac{3\,\lambda}{5}, \quad x_4 = \frac{5}{9} - \frac{3\,\lambda}{2}, \quad y_1 = \frac{7}{45} + \frac{7\,\lambda}{10}$$

On this segment of the optimal path two trades take place, namely 1-4 and 2-3. Both trades have efficiencies measured only in terms of direct benefits since $G_{14} = F_{14} = G_{23} = F_{23} = \lambda$.

By decreasing λ we find a boundary point due to constraint 2 becoming active. The breakpoint is

$$\lambda = \frac{38}{189}, \quad x_1 = \frac{26}{63}, \quad x_2 = \frac{16}{63}, \quad x_3 = \frac{5}{63} \quad x_4 = \frac{16}{63}, \quad y_1 = \frac{8}{27}$$

By computing $G_{12}(x,y) = G_{13}(x,y) = G_{14}(x,y) = \lambda$, plus the constraints, we get

$$x_1 = \frac{10}{33} + \frac{6\,\lambda}{11}, \quad x_2 = \frac{4}{11} - \frac{6\,\lambda}{11}, \quad x_3 = -\frac{1}{33} + \frac{6\,\lambda}{11}, \quad x_4 = \frac{4}{11} - \frac{6\,\lambda}{11}, \quad y_1 = \frac{14}{33} - \frac{7\,\lambda}{11}, \quad y_2 = \frac{38}{33} - \frac{63\,\lambda}{11}$$

In this case a direct measure of the trade efficiency cannot by itself justify the trades which take place on this segment of the optimal path. Consider for instance the two trades 1-2 and 3-4. In isolation the trade 1-2 cannot take place because it would oversaturate the first constraint. On the contrary the trade 3-4 could take place in isolation. It has the additional effect of desaturating the second constraint thereby making feasible the trade 1-2. This leads to an efficiency higher than that one measured simply by F_{34} and this indirect benefit is captured by H_{34}. Similarly the efficiency of the trade 1-2 is lower than it appears in F_{12} because it cannot happen without the help of another trade of lower direct efficiency, and the induced loss is indeed measured by H_{12}. Similar considerations can be applied if we consider the trades 1-4 and 2-3.

By decreasing λ we find a boundary event due to asset 3 going to zero. The breakpoint is

$$\lambda = \frac{1}{18}, \quad x_1 = \frac{1}{3}, \quad x_2 = \frac{1}{3}, \quad x_3 = 0 \quad x_4 = \frac{1}{3}, \quad y_1 = \frac{7}{18}, \quad y_2 = \frac{5}{6}$$

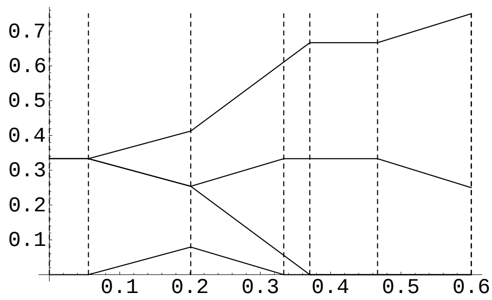
and this is also the final point in the X space. Indeed, by computing $G_{12}(x,y) = G_{14}(x,y) = \lambda$, plus the constraints, we get

$$x_1 = \frac{1}{3}, \quad x_2 = \frac{1}{3}, \quad x_3 = 0 \quad x_4 = \frac{1}{3}, \quad y_1 = \frac{2}{3} - 5\,\lambda, \quad y_2 = \frac{4}{3} - 9\,\lambda$$

and λ can go to zero without finding breakpoints. We sum up the breakpoints

λ	x_1	x_2	x_3	x_4	y_1	y_2	y_3
$\frac{3}{5}$	$\frac{3}{4}$	$\frac{1}{4}$	0	0	0	0	0
$\frac{7}{15}$	$\frac{2}{3}$	$\frac{1}{3}$	0	0	0	0	0
$\frac{10}{27}$	$\frac{2}{3}$	$\frac{1}{3}$	0	0	$\frac{13}{27}$	0	0
$\frac{1}{3}$	$\frac{11}{18}$	$\frac{1}{3}$	$\frac{1}{18}$	0	$\frac{7}{18}$	0	0
$\frac{38}{189}$	$\frac{26}{63}$	$\frac{16}{63}$	$\frac{5}{63}$	$\frac{16}{63}$	$\frac{8}{27}$	0	0
$\frac{1}{18}$	$\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{7}{18}$	$\frac{5}{6}$	0
0	$\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{4}{3}$	0

We also display in next figure the behaviour of the x_i values as a function of λ. We remark that it is only by chance that the various lines do not cross each other but they merge together. With other data the lines do cross.



Example 2: x_i as a function of λ

It may be seen that the optimum path in the extended space is made by six segments corresponding to the six intervals of λ exhibited in the figure. Moreover, there are two types of intervals: in the first type (second and last interval from right to left) only virtual assets move, while the real ones are frozen. The corresponding values of λ may not be reached in the real world, so there is no optimum point in the real world space with such level of benefit.

In the second type, on the contrary, at least some real world variables move, perhaps jointly, with virtual variables. The corresponding levels of benefit λ may then be achieved in real world. This means that in any point of the real world space, which is the projection of this part of the optimum extended path, the ratio between (one half) the directional derivative of the variance and the directional derivative of the expectation equals λ and is the biggest value of the ratio overall feasible directions.

In turn, this part of the path too may be divided into two subsets. In the first subset the efficiency associated to each point is wholly captured by the benefits (revealed by the F values) of real world bilateral trading involving at least two assets. More precisely, if we consider the movement on each segment of the real world projection of this subset, there is at least one full decomposition (segment specific) of the movement in bilateral tradings such that the efficiency of such tradings as measured by the F functions is at the common level λ. Of course λ is also the efficiency captured by the functions G for any pair of real world active assets in the segment.

In intervals of the second type, corresponding to the next-to-the-last interval (from right to left), such decompositions do not exist. This means that for each feasible decomposition there is at least one bilateral trading of the decomposition, whose F value does not match the current level of efficiency λ (revealed by G). And this happens if and only if for any feasible decomposition there is at least one pair of the decomposition whose assets are in asymmetric position (one in and one out the constraint) with respect to at least one of the binding constraints. If (and only if) this happens, we are no more able to wholly describe the optimum path in terms of direct benefits from bilateral tradings. Also indirect benefits need to be considered. The need of introducing the G functions, as a practical tool and not only as a mere elegant mathematical model, comes indeed from these cases.

10. Conclusions

It is today plainly recognized that de Finetti introduced the mean-variance approach in financial problems under uncertainty and, treating a problem of proportional reinsurance offered a correct procedure to find the mean variance efficient set in case of no correlation and under a convenient regularity assumption also in case of correlation. Furthermore, a natural extension of his procedure respecting his logic and applying his simple and original tool of the key functions, provides correct solutions to find the mean-variance optimum reinsurance set also in case of non-regularity. De Finetti did not ever discuss the applications of his approach to the asset portfolio selection, later discussed and solved by H. Markowitz by making decisive recourse to the technical tools meanwhile offered by progress in quadratic optimization.

We show in this paper that de Finetti's logic, applied through a new set of key functions describing the effects of bilateral trading reveals to be able to build a simple procedure to find the mean-variance efficient portfolios set, and, more important in our opinion, provides also a clear picture of what lies at the core of the efficiency concept.

Astonishingly simple, a portfolio is efficient if and only if any bilateral feasible "small" trading between a pair of active assets shares the same benefit, while any feasible "small" bilateral trading between one active and one non-active asset gives a lower benefit or a greater burden than the one obtained through internal trading.

Such results hold plainly in the classical case with only $n+1$ constraints of non-negativity and saturation. In order to extend the results to the case with m additional collective lower or upper bounds (threshold) constraints we need to pass to an extended $n+m$ dimensional space Z, with the addition of m virtual dimensions to the old real world n dimensional space X. A new set of extended key functions is then defined on the new space to properly capture also the indirect effect (benefit or burden) on bilateral trading induced by binding constraints.

Appendix

In the Appendix we prove some results for the case of uncorrelated risks. If there is no correlation the optimality condition (8) simplifies to

$$\frac{\sigma_{kk}\, x_k}{m_k - m_h} \geq \lambda, \quad h \in I_0^k, \qquad \frac{\sigma_{kk}\, x_k - \sigma_{hh}\, x_h}{m_k - m_h} = \lambda, \quad h \in I_k^*, \qquad \frac{\sigma_{kk}\, x_k}{m_k - m_h} \leq \lambda, \quad h \in I_k^0 \tag{25}$$

which implies $I_0^k = \emptyset$. Hence the first variable is always the reference variable and (25) is actually

$$\frac{\sigma_{11}\, x_1 - \sigma_{hh}\, x_h}{m_1 - m_h} = \lambda, \quad h \in I_k^*, \qquad \frac{\sigma_{11}\, x_1}{m_1 - m_h} \leq \lambda, \quad h \in I^0 \tag{26}$$

Note that $h \in I^0$ implies $j \in I^0$ for all $j > h$. Therefore, by decreasing λ, the variables join I^* one by one in order of decreasing values m_h and never go to zero.

Let $\hat{x}^r$ be the corner point such that the asset r (still at value 0) joins I^* and let $\hat{\lambda}^r$ the corresponding value of the efficiency index λ. By optimality we have at $\hat{x}^r$:

$$F_{ij}(\hat{x}^r) = \hat{\lambda}^r, \qquad i, j \leq r, \quad i \neq j$$

and in particular

$$F_{ir}(\hat{x}^r) = \frac{\sigma_{ii}\,\hat{x}_i^r}{m_i - m_r} = \hat{\lambda}^r, \qquad i < r$$

from which

$$\hat{x}_i^r = \hat{\lambda}^r\,\frac{m_i - m_r}{\sigma_{ii}}$$

By imposing $\sum_{i=1}^{r-1} \hat{x}_i^r = 1$, we easily get

$$\hat{\lambda}^r = \frac{1}{\sum_{j=1}^{r-1} \frac{m_j - m_r}{\sigma_{jj}}}, \qquad \hat{x}_i^r = \frac{\frac{m_i - m_r}{\sigma_{ii}}}{\sum_{j=1}^{r-1} \frac{m_j - m_r}{\sigma_{jj}}}, \quad i < r, \qquad \hat{x}_i^r = 0, \quad i \geq r$$

At the final point ($\hat{\lambda}^{n+1} = 0$) the optimality condition (25) gives

$$\sigma_{11}\,\hat{x}_1^{n+1} - \sigma_{hh}\,\hat{x}_h^{n+1} = 0$$

from which we immediately derive

$$\hat{x}_i^{n+1} = \frac{\frac{1}{\sigma_{ii}}}{\sum_{j=1}^{n} \frac{1}{\sigma_{jj}}}$$

As for the difference between two adjacent corner values of the same assets we have:

$$\hat{x}_i^{r+1} - \hat{x}_i^r = \hat{\lambda}^{r+1}\,\frac{m_i - m_{r+1}}{\sigma_{ii}} - \hat{\lambda}^r\,\frac{m_i - m_r}{\sigma_{ii}} = (\hat{\lambda}^{r+1} - \hat{\lambda}^r)\frac{m_i - m_r}{\sigma_{ii}} + \hat{\lambda}^{r+1}\,\frac{m_r - m_{r+1}}{\sigma_{ii}}$$

$$\frac{\hat{x}_i^{r+1} - \hat{x}_i^r}{\hat{\lambda}^{r+1} - \hat{\lambda}^r} = \frac{m_i - m_r}{\sigma_{ii}} + \frac{\hat{\lambda}^{r+1}}{\hat{\lambda}^{r+1} - \hat{\lambda}^r}\,\frac{m_r - m_{r+1}}{\sigma_{ii}}$$

Now

$$\frac{\hat{\lambda}^{r+1}}{\hat{\lambda}^{r+1} - \hat{\lambda}^r} = \frac{\frac{1}{\hat{\lambda}^r}}{\frac{1}{\hat{\lambda}^r} - \frac{1}{\hat{\lambda}^{r+1}}} = \frac{\sum_{j=1}^{r} \frac{m_j - m_r}{\sigma_{jj}}}{\sum_{j=1}^{r} \frac{m_j - m_r}{\sigma_{jj}} - \sum_{j=1}^{r} \frac{m_j - m_{r+1}}{\sigma_{jj}}} =$$

$$\frac{\sum_{j=1}^{r} \frac{m_j - m_r}{\sigma_{jj}}}{\sum_{j=1}^{r} \frac{m_{r+1} - m_r}{\sigma_{jj}}} = \frac{\sum_{j=1}^{r} \frac{m_j - m_r}{\sigma_{jj}}}{(m_{r+1} - m_r)\sum_{j=1}^{r} \frac{1}{\sigma_{jj}}}$$

so that

$$\frac{\hat{x}_i^{r+1} - \hat{x}_i^r}{\hat{\lambda}^{r+1} - \hat{\lambda}^r} = \frac{m_i - m_r}{\sigma_{ii}} - \frac{\sum_{j=1}^{r} \frac{m_j - m_r}{\sigma_{jj}}}{\sigma_{ii}\sum_{j=1}^{r} \frac{1}{\sigma_{jj}}} = \frac{\sum_{j=1}^{r} \frac{m_i - m_r}{\sigma_{jj}} - \sum_{j=1}^{r} \frac{m_j - m_r}{\sigma_{jj}}}{\sigma_{ii}\sum_{j=1}^{r} \frac{1}{\sigma_{jj}}} = \frac{\sum_{j=1}^{r} \frac{m_i - m_j}{\sigma_{jj}}}{\sigma_{ii}\sum_{j=1}^{r} \frac{1}{\sigma_{jj}}} = \frac{\Phi_i^r}{\sigma_{ii}\sum_{j=1}^{r} \frac{1}{\sigma_{jj}}}$$

References

de Finetti, B. (1940), "Il problema dei pieni", *Giornale Istituto Italiano Attuari*, **9**, 1-88 ; English translation by L. Barone available as "The problem of "Full-risk insurances", Ch. 1 'The problem in a single accounting period', *Journal of Investment Management*, **4**, 19-43, 2006.

Karush, W. (1939), *Minima of functions of several variables with inequalities as side constraints*, M.Sc. dissertation, Department of Mathematics, University of Chicago, Chicago, IL, USA.

KUHN, H.W. AND A.W. TUCKER (1951), "Nonlinear programming", in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman ed., University of California Press, Berkeley, CA, USA.

LINTNER, J. (1965), "The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets", *Review of Economic and Statistics*, **47**, 13-37.

MARKOWITZ, H. (1952), "Portfolio selection", *Journal of Finance*, **6**, 77-91.

MARKOWITZ, H. (1956), "The optimization of quadratic functions subject to linear constraints", *Naval Research Logistics Quarterly*, **3**, 111-133.

MARKOWITZ, H.(2006), "de Finetti scoops Markowitz", *Journal of Investment Management*, **4**, 5-18.

MITRA G., T. KYRIAKIS, C. LUCAS AND M. PIRBHAI (2003), "A Review of Portfolio Planning: Models and Systems", in *Advances in Portfolio Construction and Implementation*, S.E. Satchell and A.E. Scowcroft (eds.), Butterworth and Heinemann, Oxford, UK.

PRESSACCO, F. (1986), "Separation theorems in proportional reinsurance", in *Insurance and Risk Theory*, M. Goovaerts et al. eds., Reidel Publishing, pp. 209-215.

PRESSACCO, F. AND P. SERAFINI (2007), "The origins of the mean-variance approach in finance: revisiting de Finetti 65 years later", *J. of Decisions in Economics and Finance*, **30**, 19-49.

RUBINSTEIN M. (2006), "Bruno de Finetti and mean-variance portfolio selection", *Journal of Investment Management*, **4**, 3-4.

SHAPIRO, J.F. (1979), *Mathematical programming: structures and algorithms*, Wiley, New York, NY, USA.

TOBIN, J. (1958), "Liquidity preference as behaviour toward risk", *Review of Economic Studies*, **36**, 65-86.