A global optimization problem in portfolio selection

M.C. Bartholomew-Biggs · **S.J. Kane**

Abstract This paper deals with the issue of buy-in thresholds in portfolio optimization using the Markowitz approach. Optimal values of invested fractions calculated using, for instance, the classical minimum-risk problem can be unsatisfactory in practice because they lead to unrealistically small holdings of certain assets. Hence we may want to impose a discrete restriction on each invested fraction y_i such as $y_i > y_{min}$ or $y_i = 0$. We shall describe an approach which uses a combination of local and global optimization to determine satisfactory solutions. The approach could also be applied to other discrete conditions - for instance when assets can only be purchased in units of a certain size (roundlots).

Keywords Portfolio selection · buy-in thresholds · roundlots · global optimization · mixed-integer optimization

---

The authors are grateful to the anonymous referees for comments and suggestions which have helped to strengthen this paper. Particular thanks are due to Don Jones for generous advice about our discussion and use of DIRECT.

M.C. Bartholomew-Biggs · S.J. Kane
School of Physics Astronomy and Mathematics
University of Hertfordshire
Hatfield AL10 9AB
England

1 Introduction

In this paper we consider the standard approach to portfolio optimization proposed by Markowitz (1952,1959). We suppose that, for a set of n assets, we have the mean returns $\bar{r}_1,...,\bar{r}_n$ and the $n \times n$ variance-covariance matrix Q based on a past performance history. If a portfolio is defined by the invested fractions $y_1,...,y_n$ then its expected return, R, and its risk V are denoted by

$$R = \bar{r}^T y \quad \text{and} \quad V = y^T Q y. \tag{1}$$

Initially, we shall require the invested fractions to satisfy

$$\sum_{i=1}^{n} y_i = 1 \quad \text{(or equivalently } \; e^T y = 1 \quad \text{where } \; e^T = (1,1,...,1)\ ). \tag{2}$$

In its simplest form, the problem of choosing invested fractions to obtain a minimum risk portfolio with expected return $R = R_p$ can be expressed as a quadratic programming problem

$$\text{Minimize} \quad y^T Q y \quad \text{s.t.} \;\; r^T y = R_p \;\; \text{and} \;\; e^T y = 1. \tag{3}$$

If we wish to exclude the possibility of short-selling we must add inequality constraints as in

$$\text{Minimize} \quad y^T Q y \quad \text{s.t.} \;\; \bar{r}^T y = R_p, \;\; e^T y = 1 \quad \text{and} \;\; y_i \geq 0, \; i = 1,..,n. \tag{4}$$

In practice we might want to put a non-zero lower bound, y_{min}, on each y_i either to avoid transaction costs on very small trades or simply because a seller imposes some minimum threshold on the purchase of each asset. However, if we simply replace the inequalities in (4) by $y_i \geq y_{min}$ for $i = 1,...,n$ then we exclude the possibility that making no investment in an asset might be better than acquiring some (possibly rather arbitrary) minimum holding. The constraint we would like to apply instead is one of the form

$$\text{either} \;\; y_i = 0 \quad \text{or} \quad y_i \geq y_{min}. \tag{5}$$

This kind of *buy-in threshold* constraint is discussed, for instance, by Jobst et. al. (2001) and Mitra et. al. (2003). A standard way of dealing with the minimum-risk problem with buy-in thresholds is by finding $y_1,...,y_n$ and $z_1,...,z_n$ to solve a mixed integer quadratic programming problem (MIQP) such as

$$\text{Minimize} \quad y^T Q y \quad \text{s.t.} \;\; \bar{r}^T y = R_p, \;\; e^T y = 1 \quad \text{and} \;\; y_i \geq z_i y_{min}, \; i = 1,..,n \tag{6}$$

with additional constraints, for $i = 1, .., n$,

$$1 \geq z_i \geq 0 \quad \text{and} \quad z_i \quad \text{is integer.} \tag{7}$$

Branch-and-bound methods are typically used to solve problems like (6), (7) as discussed, for instance, by Mitra et. al. (2003). We shall say a little more about this in section 5; but our main purpose in this paper is to demonstrate the possibility of solving problems with buy-in threshold constraints by means of a global minimization approach which involves continuous rather than discrete variables. We introduce the function

$$\phi(y_i) = 4\frac{y_i(y_i - y_{min})}{y_{min}^2}. \tag{8}$$

This function is non-negative when $y_i \leq 0$ or $y_i \geq y_{min}$ and $0 > \phi(y_i) \geq -1$ when y_i is in the unacceptable range between y_{min} and zero. Hence we can replace (6), (7) by the problem

$$\text{Minimize} \quad y^T Q y \quad \text{s.t.} \ r^T y = R_p, \ e^T y = 1 \quad \text{and} \ y_i \geq 0, \ i = 1, .., n \tag{9}$$

with the additional nonlinear (and nonconvex) constraints

$$\phi(y_i) \geq 0, \quad \text{for} \ i = 1, ..., n. \tag{10}$$

We can expect (9), (10) to have several local solutions at which some of the constraints (10) will be binding with either $y_i = 0$ or $y_i = y_{min}$. As a first attempt at investigating global solutions of this problem we shall use a penalty function approach. Our reason for adopting this somewhat old-fashioned technique is that it enables us to apply the DIRECT method described by Jones et. al. (1993) and Jones (2001). This is an algorithm for unconstrained[1] global minimization that we have used successfully in other applications (Bartholomew-Biggs et. al. (2003, 2005)).

In order to construct a penalty function corresponding to (9), (10) we shall handle the positivity constraints on y_i by introducing new optimization variables $x_1, ..., x_n$ and then using $y_i = x_i^2, \ i = 1, ..., n$ in all subsequent expressions. The penalty function that we seek to minimize is then

$$F = y^T Q y + \rho(e^T y - 1)^2 + \rho(\frac{\bar{r}^T y}{R_p} - 1)^2 + \mu \sum_{i=1}^{n} \psi(y_i)^2 \tag{11}$$

[1]DIRECT cannot deal with problems with equality constraints. There *is* however a version which handles problems with inequality constraints and we shall have a little more to say about this in a later section.

where

$$\psi(y_i) = \text{Min}(0, \phi(y_i)), \quad i = 1, ..., n \tag{12}$$

and where ρ and μ are suitably large positive penalty parameters.

We can apply a similar approach to the maximum return problem

$$\text{Minimize } -\bar{r}^T y \quad \text{s.t. } y^T Q y = V_a \text{ and } e^T y = 1 \tag{13}$$

where V_a denotes an acceptable level of portfolio risk. If we wish to exclude short-selling and also include buy-in threshold constraints then we can do so by treating $x_1, ..., x_n$ as variables, setting $y_i = x_i^2$ and minimizing the penalty function

$$F = -\bar{r}^T y + \rho(e^T y - 1)^2 + \rho(\frac{y^T Q y}{V_a} - 1)^2 + \mu \sum_{i=1}^{n} \psi(y_i)^2 \tag{14}$$

where $\psi(y_i)$ is given by (12).

The next section gives a brief outline of DIRECT and then, in section 3, we present some numerical results obtained by applying DIRECT to (11) and (14). Section 4 introduces a variant of (11) to deal with the *roundlot* problem. In section 5, we discuss other approaches to similar portfolio selection problems and we also identify topics for further investigation.

2 DIRECT

In practice, most methods which seek the global minimum of a function $F(x)$ are applied in some restricted region of variable-space, typically in a “hyperbox”

$$l_i \le x_i \le u_i.$$

The algorithm DIRECT (Jones et al (1993), Jones (2001)) works by systematic division of the original search region into smaller sub-boxes. In the limit, as the number of iterations becomes infinite, it will sample the whole region and, in that sense, the algorithm is guaranteed to converge. DIRECT performs well in practice because of the way it chooses which sub-boxes to explore first. This frequently enables it to get a reasonable approximation to a global minimum in a fairly small number of iterations (after which a more rapidly convergent local minimization method can be used to obtain the solution to greater accuracy).

We begin by outlining how DIRECT works for a one-variable problem and then we show how the ideas can be extended to two or more variables. Consider the problem of finding the global minimum of $F(x)$ for $0 \leq x \leq 1$. Initially, DIRECT divides the range $[0,1]$ into three equal parts and evaluates the function at their midpoints. The sub-range which has the least function value is then trisected and F is calculated at the centre point of all the new ranges. We then have a situation like that shown in Figure 1.

```
                              F

                                                           F
          F
                                                    F
                                            F

0___________________|___________________|_______|_______|______1
```

Figure 1: **One iteration of DIRECT on a one-variable problem**

There are now trial ranges of two different widths, namely, $\frac{1}{3}$ and $\frac{1}{9}$. For *each* of these widths, DIRECT considers the one with the smallest value of F at the centre as a candidate for further subdivision. If we suppose for the moment that both candidates are trisected then the outcome could be as depicted in Figure 2.

```
                              F

                                                                F
           F
                   F                                    F
    F                                         F   F
                                          F
0_____|_______|_______|_________________|___|___|___|_______|______1
A     B       C       D                 E   F   G   H       I      J
```

Figure 2: **Two iterations of DIRECT on a one-variable problem**

After this second iteration there are *three* candidate range-sizes, $\frac{1}{3}$, $\frac{1}{9}$ and $\frac{1}{27}$; and for each of these, the one with smallest central F-value is considered for further subdivision. In Figure 2, the candidate intervals would be DE, AB and EF. Continuation of this process amounts to a systematic exploration the whole range, giving priority to the most promising regions.

The distinctive and effective feature of DIRECT is the way it identifies which of the candidate ranges are "promising". Suppose that $d_1, .., d_p$ are the p different range-sizes at the start of an iteration. Suppose also that F_j denotes the smallest of the mid-point function values in ranges of width d_j. DIRECT will trisect the range containing F_j *only* if a "potential optimality" test is satisfied. This test is based upon *Lipschitz constants* (i.e. bounds on the magnitude of the first derivative of F) and is described fully by Jones et. al. (1993). Briefly, the test is aimed at avoiding any waste of computing effort on the subdivision of ranges which *on present evidence* are unlikely to contain the global optimum. The potential optimality test discourages unneccessarily close examination of ranges containing local minima but ensures that large unexplored ranges can always be revisited.

The ideas outlined for a one-variable problem are the basis of DIRECT for problems in two or more variables as first proposed by Jones et. al. (1993). For a two-variable problem, the original search region is now a rectangle rather than a line segment[2]. The initial subdivision is into three sub-boxes by trisection along the longest edge. The objective function is evaluated at the centre of each sub-box and the size of a sub-box is taken as the length of its diagonal. The box with the smallest central value of F is subdivided by trisection along its longest side[3] and the process of identification and subdivision of potentially optimal boxes then continues as in the one-variable case.

The further extension of DIRECT to deal with problems in n variables can easily be inferred from the above discussion of the one and two dimensional cases simply by rephrasing the previous paragraph in terms of hyper-rectangles and hyperboxes.

Experience in a number of practical situations has shown that DIRECT can get good estimates of global optima quite quickly. Since it only uses function values, it can be applied to non-smooth problems or to those where the computation of derivatives is difficult. One drawback, however, is that there is no hard-and-fast convergence test for stopping the algorithm. One can simply let it run for a fixed number of iterations or else choose to terminate if there is no improvement in the best function value after a prescribed number of function evaluations. Neither of these strategies, however, will *guarantee* that enough iterations have been done to identify the neighbourhood of the global optimum.

It should be emphasised, of course, that many other global optimization algorithms exist and the implementation of minimization methods involving (11) and (14) does not rely exclusively on the use of any particular method.

---

[2]By a suitable change of variables this rectangle can conveniently be made into a unit square

[3]Refinements for subdividing boxes with several longest sides are given by Jones et. al. (1993).

3 Numerical results

We now give some demonstration examples involving small portfolios. Specifically, for a group of five real-life assets we use historical stock market data to generate mean returns $\bar{r}$ and a variance-covariance matrix Q. We then solve problem (4) (or problem (13) with the addition of positivity constraints) and, on the basis of these computed solutions, we specify a value for y_{min} and seek a modified portfolio to satisfy the buy-in threshold constraints (5). We do this by seeking global minima of the penalty functions (11) or (14).

As mentioned above, (11) and (14) are actually treated as functions of artificial variables x_i with the invested fractions being taken as $y_i = x_i^2$ to prevent solutions involving short-selling. In the interests of clarity, however, the descriptions of numerical results will be expressed entirely in terms of the y_i variables.

3.1 Minimum-risk solutions with buy-in threshold

For the five assets in our sample problem, the vector of mean returns is

$$\bar{r} = (-0.056,\ 0.324,\ 0.343,\ 0.132,\ 0.108)^T.$$

The variance-covariance matrix is

$$Q \approx \begin{pmatrix} 2.4037 & -0.0222 & 0.5230 & 0.2612 & 0.6126 \\ -0.0222 & 1.8912 & 0.0442 & 0.0020 & 0.4272 \\ 0.5230 & 0.0442 & 1.7704 & 0.2283 & 0.3103 \\ 0.2612 & 0.0020 & 0.2283 & 4.4812 & -0.1134 \\ 0.6126 & 0.4272 & 0.3103 & -0.1134 & 7.7490 \end{pmatrix}.$$

The values in $\bar{r}$ imply that a reasonable choice of target return is $R_p = 0.25\%$. The corresponding solution to (4) has invested fractions

$$y_1 \approx 0.132,\ y_2 \approx 0.368,\ y_3 \approx 0.345,\ y_4 \approx 0.117,\ y_5 \approx 0.037 \qquad (15)$$

giving a portfolio risk $V \approx 0.6894$.

We note that (15) includes a relatively small investment in asset five and so we consider (11) with $y_{min} = 0.05$ and penalty parameters $\rho = 10^3$ and $\mu = 1$. At the minimum of (11) we expect *either* that y_5 will be near zero *or* that $y_5 \approx 0.05$. That is, we expect a change of about ± 0.03 in y_5. In order to maintain the total investment $\sum y_i = 1$, this means there could be a compensating change of up to

± 0.03 in any of the other invested fractions. Therefore we shall seek the *global* optimum of (11) in the hyperbox defined by

$$0.1 \leq y_1 \leq 0.16;\ 0.34 \leq y_2 \leq 0.4;\ 0.31 \leq y_3 \leq 0.37; \quad (16)$$

$$0.09 \leq y_4 \leq 0.15;\ 0 \leq y_5 \leq 0.06. \quad (17)$$

We have used intuitive arguments about the preliminary solution (15) to set quite tight bounds on the invested fractions. This is worth doing as it can be expected to speed up convergence to the global minimum of (11). However it should be understood that the problem could still be solved if wider bounds than (16), (17) were set.

Before using DIRECT it is interesting to apply a standard quasi-Newton *local* minimization method to (11). Specifically we use a BFGS algorithm (Broyden, 1970a, 1970b). If we start from the midpoint of the region (16), (17), where (11) has a value of about 2.49, the quasi-Newton method converges to a minimum with

$$y_1 \approx 0.152,\ y_2 \approx 0.381,\ y_3 \approx 0.348,\ y_4 \approx 0.118,\ y_5 \approx 0 \quad (18)$$

where (11) has a value approximately 0.7005. The portfolio risk is 0.7001 and the return is 0.25%, as required.

We now apply DIRECT to (11) within the same hyperbox (16), (17). After 10 iterations it gives a point where (11) has a value of about 0.6949. This is already appreciably better than the function value at (18) which implies that the optimum found by the quasi-Newton method is only a local solution. The invested fractions given by DIRECT after 20 iterations are

$$y_1 \approx 0.122,\ y_2 \approx 0.369,\ y_3 \approx 0.339,\ y_4 \approx 0.111,\ y_5 \approx 0.058 \quad (19)$$

with a portfolio risk of 0.6935. This approximate solution has $y_5 \approx 0.05$ in contrast to (18), which has $y_5 \approx 0$.

Although DIRECT has fairly easily obtained a better solution than the quasi-Newton method, it may not be very efficient at finding optima to high accuracy. This is partly because it does not use derivatives and partly because it only samples function values at the centres of hyperboxes. A common strategy, therefore, is to run a quasi-Newton method from the best point located by DIRECT, in order to refine the approximate solution. If we apply this strategy to the estimate (19), we find that an accurate global minimum of (11) has a value of about 0.691 and involves the invested fractions

$$y_1 \approx 0.125,\ y_2 \approx 0.364,\ y_3 \approx 0.344,\ y_4 \approx 0.116,\ y_5 \approx 0.05. \quad (20)$$

Here the target return 0.25% is still achieved and the portfolio risk is 0.6906. This is only slightly worse than the risk for the portfolio (15) which was obtained without considering the buy-in threshold constraint ($y_i = 0$ or $y_i \geq 0.05$).

3.2 Maximum return with buy-in threshold

For this example we use the same dataset as in the previous section and take the acceptable risk as $V_a = 0.75$ which is slightly higher than the minimum risk associated with an expected return of 0.25%. The solution of the maximum return problem (13) without short-selling is

$$y_1 \approx 0.0462,\ y_2 \approx 0.4022,\ y_3 \approx 0.4162,\ y_4 \approx 0.1053,\ y_5 \approx 0.0299 \qquad (21)$$

giving a portfolio return of 0.2877%.

To obtain a portfolio in which all non-zero invested fractions are greater than or equal to 0.05 we minimize (14) with $y_{min} = 0.05, \rho = 10^3$ and $\mu = 1$. A reasonable box to search in is

$$0 \leq y_1 \leq 0.06;\ 0.36 \leq y_2 \leq 0.44;\ 0.38 \leq y_3 \leq 0.46; \qquad (22)$$

$$0.06 \leq y_4 \leq 0.12;\ 0 \leq y_5 \leq 0.06. \qquad (23)$$

When started from the midpoint of the region (22), (23), the quasi-Newton method finds a minimum at

$$y_1 \approx 0.0034,\ y_2 \approx 0.4037,\ y_3 \approx 0.4112,\ y_4 \approx 0.1529,\ y_5 \approx 0.0030 \qquad (24)$$

where (14) has a value of about 0.6472. Twenty iterations of DIRECT in the same box, however, produce the much lower value, $F = -0.276$. This occurs when

$$y_1 \approx 0.0578,\ y_2 \approx 0.399,\ y_3 \approx 0.419,\ y_4 \approx 0.0746,\ y_5 \approx 0.0494. \qquad (25)$$

Hence it is better for y_1 and y_5 to be near 0.05 rather than near zero, as they are in the *local* solution (24) produced by the quasi-Newton method.

If we use the quasi-Newton method to refine the approximate global minimum (25) we get the more accurate result

$$y_1 = 0.05,\ y_2 \approx 0.396,\ y_3 \approx 0.417,\ y_4 \approx 0.0869,\ y_5 = 0.05. \qquad (26)$$

Here the portfolio return is about 0.2855%, which is only slightly worse than was possible when (5) is not enforced.

3.3 Discussion and extensions

The two previous examples provide *prima facie* evidence that global minimization of the penalty functions (11) and (14) can be used to solve portfolio optimization problems involving disjoint constraints like (5). To strengthen that evidence, we first point out that the success of DIRECT as a global minimizer of (11) does not depend on us restricting the search to a small hyperbox such as (16), (17). For instance, if we apply DIRECT to (11) using the same data as in section 3.1 but with the much larger search region

$$0 \leq y_i \leq 0.5, \quad \text{for} \ \ i = 1, .., 5 \tag{27}$$

then, after about 100 iterations, we obtain a point similar to (19). This can be refined to the accurate solution (20) by a few quasi-Newton iterations. Similarly, the maximum-return problem in section 3.2 can also be solved by applying DIRECT to (14) in the hyperbox (27) rather than (22), (23). This takes about 150 iterations of DIRECT followed by quasi-Newton refinement.

We now describe a more general way of using (11) for problems with larger numbers of assets. In spite of the comments in the preceding paragraph, the approach we propose does make use of a preliminary solution of (4), as outlined below.

Find values $\hat{y}_i$ to solve (4).
Obtain new trial values $\tilde{y}_i$ by

$$\tilde{y}_i = \begin{cases} 0 & \text{if } \hat{y}_i = 0 \\ y_{min} & \text{if } 0 < \hat{y}_i \leq y_{min} \\ \hat{y}_i & \text{if } \hat{y}_i > y_{min} \end{cases}$$

Minimize (11) *using* $y_1, .., y_n$ *as variables* by applying DIRECT in the hyperbox

$$\begin{array}{ll} y_i = 0 & \text{if } \tilde{y}_i = 0 \\ \tilde{y}_i - y_{min} \leq y_i \leq \tilde{y}_i + y_{min} & \text{if } \tilde{y}_i > 0 \end{array}$$

Minimize (11) *using* $x_1, \ldots, x_n$ *as variables* by a quasi-Newton method (starting from the solution obtained by DIRECT)

We can avoid the $y_i = x_i^2$ transformation when minimizing (11) with DIRECT because the hyperbox limits ensure that short-selling will not occur. However the final refinement should be done in terms of the artificial variables x_i because the quasi-Newton method does not put any restrictions on the y_i.

Table 1 shows the results of applying the strategy to a ten-asset problem, using $y_{min} = 0.05$. The final section of Table 1 shows that the constraint (5) can be satisfied for a relatively small increase in risk. We emphasise again that the use of a *global* optimizer like DIRECT is important. If we had attempted to solve the problem simply by applying a quasi-Newton method to (11) using $x_i = \sqrt{\tilde{y}_i}$ as a starting point then we would only have obtained a local solution with y_1, y_4, y_5 and y_7 all at the limiting value 0.05 and a substantially higher risk value of 0.411.

Table 1: **Ten asset problem with buy-in threshold constraint**

Solution from (4) $y \approx (0.054, 0.227, 0.185, 0.055, 0.035, 0.098, 0.007, 0.099, 0.2, 0.041)$ Risk$\approx$ 0.384
Minimizing (11) with $\rho = 1000, \mu = 1$ (using DIRECT) $y \approx (0.071, 0.227, 0.180, 0.055, 0.05, 0.066, 0.0, 0.104, 0.197, 0.05)$ Risk$\approx$ 0.392
Minimizing (11) with $\rho = 1000,\ \mu = 1$ (using DIRECT + q-N refinement) $y \approx (0.058, 0.215, 0.173, 0.06, 0.05, 0.096, 0.0, 0.098, 0.2, 0.05)$ Risk$\approx$ 0.387

As a final example we consider a 50 asset problem with $y_{min} = 0.03$. Calculated portfolios to give minimum risk for an expected return of 0.1% are summarised in Table 2. The entries in the table show how the distribution of non-zero invested fractions changes as the threshold constraint (5) is taken into account.

Table 2: **Fifty asset problem with buy-in threshold constraint**

	Assets with $0 < y_i < 0.03$	Assets with $y_i = 0.03$	Assets with $y_i > 0.03$
Solution from (4) Risk $\approx$ 0.744 & 50	1,8,15, 34,44,47	-	12,17,18, 19,20,31
Minimizing (11) (q-N only) with $\rho = 5000, \mu = 1$ Risk $\approx$ 0.778	-	8,12,15, 34,44,50	1,17,19,20, 31,47
Minimizing (11) (DIRECT) with $\rho = 5000, \mu = 1$ Risk $\approx$ 0.764	-	47, 50	1,8,12,15, 17,18,19,20, 31
Minimizing (11) (DIRECT + q-N) with $\rho = 5000, \mu = 1$ Risk $\approx$ 0.757	-	15,47	1,8,12,17, 18,19,20,31, 50

In the results in Table 2 we observe how the smaller investments in the solution of (4) are re-allocated when we consider the extended function (11). When (11) is minimized by the quasi-Newton method we only get a local solution with six of the y_i fixed at the threshold value 0.03. Fifty iterations of DIRECT yield an improved portfolio with only two of the y_i on the non-zero boundary. From here, a further local refinement gives the still better result in the last section of the table.

The approach introduced in this section will be discussed further in section 5. Before that, however, we briefly consider the application of a similar strategy to another variant of the portfolio selection problem.

4 Roundlot constraints

The invested fractions y_i obtained at the solution of a minimum-risk problem must, in practice, be converted to actual numbers of shares, bonds etc that are to be purchased. If the total investment is M and if the price of asset i is p_i then the number of items in the holding of asset i should be

$$a_i = \frac{My_i}{p_i}. \tag{28}$$

Obviously a_i must be an integer: and, more likely, it must be a multiple of some *lot size* such as 10 or 100. This will mean that we will have to round the values of y_i obtained by solving (4). In effect we want to apply a constraint

$$a_i \text{ is a multiple of some integer lot size } L_i. \tag{29}$$

If we define

$$\theta(y_i) = \frac{My_i}{p_i} - \lfloor \frac{My_i}{p_i} \rfloor, \tag{30}$$

where $\lfloor v \rfloor$ denotes the integer part of a real value v, then we want the invested fractions to satisfy the constraints

$$\phi(y_i) = \theta(y_i)(1 - \theta(y_i)) = 0 \quad \text{for} \quad i = 1,..,n. \tag{31}$$

Therefore, following the ideas proposed above, we could consider solving the minimum risk problem with roundlot constraints by minimizing

$$y^T Qy + \rho(e^T y - 1)^2 + \rho(\frac{\bar{r}^T y}{R_p} - 1)^2 + \mu \sum_{i=1}^{n} \phi(y_i)^2. \tag{32}$$

However, if the optimal y_i are to be adjusted to satisfy the roundlot constraints then the condition $e^T y = 1$ may not hold precisely – i.e., we may not be able to convert all of our investment into assets. We can only require that $e^T y \leq 1$; and this means that we must use another definition of risk (see, for instance, Mitchell & Braun, 2002), namely

$$V = \frac{y^T Qy}{(e^T y)^2}. \tag{33}$$

Therefore a penalty function for the roundlot constrained problem is

$$\frac{y^T Qy}{(e^T y)^2} + \rho[Min(0, 1 - e^T y)]^2 + \rho(\frac{\bar{r}^T y}{R_p} - 1)^2 + \mu \sum_{i=1}^{n} \phi(y_i)^2. \tag{34}$$

We can now approach the minimum risk problem with constraint (31) as follows.

Find values $\hat{y}_i$ to solve (4).
Obtain new trial values $\tilde{y}_i$ from

$$\tilde{y}_i = \begin{cases} 0 & \text{if } \hat{y}_i = 0 \\ L_i & \text{if } 0 < \hat{y}_i \leq L_i \\ \hat{y}_i & \text{if } \hat{y}_i > L_i \end{cases}$$

Solve (34) *using* $y_1, .., y_n$ *as variables* by applying DIRECT in the hyperbox

$$\begin{array}{ll} y_i = 0 & \text{if } \tilde{y}_i = 0 \\ \tilde{y}_i - L_i \leq y_i \leq \tilde{y}_i + L_i & \text{if } \tilde{y}_i > 0 \end{array}$$

We do not use a quasi-Newton minimization of (34) as a refinement step because $\phi(y_i)$ depends on $\theta(y_i)$ which is non-differentiable.

As an illustrative example we consider the five- and ten-asset problems from sections 3.1 and 3.3 respectively. For simplicity we shall take the values $M = 1000$ and $p_i = 1,\ L_i = 10$ in (28), (29) for $i = 1, .., n$. This means that, ideally, we want the y_i which minimize (34) to have zeros in the third and subsequent decimal places. Some solutions are given in Tables 3 and 4, both of which compare portfolios calculated with and without roundlot constraints.

In each case, the inclusion of constraint (31) only produces a small increase in risk compared with the solution of (4). It is interesting to note, however, that the portfolios obtained using (34) are *not* what would be obtained simply by rounding the invested fractions from (4) to the nearest multiple of 10. Indeed, for these two examples, if the y_i were obtained by such nearest rounding they would not be acceptable because they would not give $e^T y \leq 1$ and because the expected portfolio return would not be the target value 0.25%. Therefore the minimization of (34) offers a reasonable basis for determining the best way to obtain a practical solution from the invested fractions which solve (4).

Table 3: **Five asset problem with roundlot constraint**

Solution from (4)
$y \approx (0.1319, 0.3686, 0.3452, 0.1168, 0.0374)$
$\sum y_i = 1$, Risk ≈ 0.69
Minimizing (34) with $\rho = 1000,\ \mu = 1$
$y \approx (0.140, 0.370, 0.350, 0.110, 0.030)$
$\sum y_i = 1$, Risk ≈ 0.6908

Table 4: **Ten asset problem with roundlot constraint**

Solution from (4)
$y \approx (0.054, 0.227, 0.185, 0.055, 0.035, 0.098, 0.007, 0.099, 0.2, 0.041)$
$\sum y_i = 1$, Risk ≈ 0.3843
Minimizing (34) with $\rho = 1000,\ \mu = 1$
$y \approx (0.050, 0.230, 0.190, 0.050, 0.030, 0.10, 0.010, 0.090, 0.20, 0.040)$
$\sum y_i = 0.99$, Risk ≈ 0.3847

5 Discussion and conclusions

In this paper we have suggested a way of handling disjoint constraints (buy-in thresholds and roundlot constraints) occurring in portfolio selection problems. The approach, which is justified by some numerical experience, involves the introduction of nonconvex constraints – defined by (8), (10) for buy-in thresholds and by (30), (31) for roundlot constraints. We have then solved the resulting nonlinear programming problems (e.g. (9), (10)) by applying an unconstrained global optimization procedure to the penalty functions (11), (14) and (34). We have chosen to use DIRECT (Jones et. al., 1993) as the global optimization procedure, but other methods could be used instead. DIRECT does have the advantage of being a non-gradient method, which is useful for the problem in section 4 where the formulation of the roundlot constraint involves a non-differentiable function (30).

The solution techniques proposed in this paper are essentially prototypes and, although the results reported in sections 3 and 4 are quite promising, there is still scope for further investigation and improvement. We present a discussion below which "works outwards" from quite specific points about the algorithmic steps in sections 3 and 4 to a broader critique of the approach itself and finally a consideration of other ways that have been (or could be) used to approach the problems we are concerned with.

We first of all acknowledge that the algorithms outlined in sections 3 and 4 are simple and rather pragmatic. The starting guess and box-size for applying DIRECT to a penalty function is generated from a solution of (4) on the basis of two plausible but unchecked assumptions. These are
(a) Any assets excluded from the portfolio given by (4) will not figure in the solution which takes account of buy-in thresholds.
(b) To accomodate the constraints (5), the changes to non-zero invested fractions $\hat{y}_i$ will be confined to the range $[-y_{min}, y_{min}]$.
On the basis of (limited) numerical tests, these assumptions seem to be justified when y_{min} is small. Further investigation may show a need for a more sophisticated way of providing initial conditions for the global minimization.

We cannot claim that use of penalty functions like (11), (14) or (34) is the best way to solve a nonlinear programming problem. Since they feature the classical squared penalty term for violated constraints, their minima are only approximations to the true constrained solutions which become exact only as ρ and μ approach infinity. For our purposes in this paper, the solutions we have obtained using fixed values of ρ and μ have been adequate for demonstrating the viability of our approach. But it might be more appropriate to use exact penalty functions obtained by replacing the squared-penalty terms by absolute values, as in

$$F = y^T Qy + \rho|e^T y - 1| + \rho|\frac{\bar{r}^T y}{R_p} - 1| + \mu \sum_{i=1}^{n} |\psi(y_i)|. \qquad (35)$$

For ρ and μ "sufficiently large", the minimum of (35) coincides with the constrained solution of the minimum risk problem. We note that DIRECT would still be a suitable algorithm for seeking the global minimum of this non-differentiable function.

Our underlying reason for using a penalty function approach was that DIRECT was originally proposed by Jones et.al. (1993) as a method for unconstrained global optimization. However, a version of DIRECT for inequality constrained problems has also been developed (Jones, 2001) and this could be an alternative method for nonlinear programming problems arising in minimum-risk portfolio selection with buy-in constraints. We cannot apply DIRECT to problem (9), (10) because the method cannot handle equality constraints; but we can apply it to the modified formulation

$$\text{Minimize} \quad \frac{y^T Qy}{(e^T y)^2} \quad \text{s.t.} \ \bar{r}^T y \geq R_p, \ e^T y \leq 1 \quad \text{and} \quad y_i \geq 0, \ i = 1,..,n \qquad (36)$$

with the additional nonlinear constraints

$$\phi(y_i) \geq 0, \text{ for } i = 1, \dots, n. \tag{37}$$

This problem is less restrictive than (9), (10) because it allows portfolio return to exceed the target R_p if this would imply a decrease in risk. Note also that this problem allows $e^T y < 1$ – i.e. we might keep some of our capital as cash – and so the risk is defined by (33).

The version of DIRECT for inequality constraints, described by Jones (2001), can also handle integer variables and so it can be applied to an inequality constrained version of (6), (7), namely

$$\text{Minimize} \quad \frac{y^T Q y}{(e^T y)^2} \text{ s.t. } \bar{r}^T y \geq R_p, \ e^T y \leq 1 \text{ and } y_i \geq z_i y_{min}, \ i = 1, .., n \tag{38}$$

with additional constraints

$$1 \geq z_i \geq 0 \text{ and } z_i \text{ is integer for } i = 1, .., n. \tag{39}$$

In this problem, the highly nonlinear constraint (37) is avoided but at the cost of introducing additional integer variables $z_1, \dots, z_n$. It remains a matter for further investigation whether (38), (39) is preferable to (36), (37) – and whether either or both are superior to the penalty function approach used in this paper.

To handle the maximum-return problem with buy-in threshold constraints we could apply the appropriate version of DIRECT to one of

$$\text{Minimize} \quad -\bar{r}^T y \text{ s.t.} \frac{y^T Q y}{(e^T y)^2} \leq V_a, \ e^T y \leq 1 \text{ and } y_i \geq 0, \ i = 1, .., n \tag{40}$$

$$\text{and } \phi(y_i) \geq 0, \text{ for } i = 1, \dots, n. \tag{41}$$

or, introducing extra variables $z_1, \dots, z_n$,

$$\text{Minimize} \quad -\bar{r}^T y \text{ s.t. } \frac{y^T Q y}{(e^T y)^2} \leq V_a, \ e^T y \leq 1 \text{ and } y_i \geq z_i y_{min}, \ i = 1, .., n \tag{42}$$

$$\text{and } 1 \geq z_i \geq 0 \text{ and } z_i \text{ is integer for } i = 1, \dots, n. \tag{43}$$

A nonlinear programming form of the minimum-risk problem with roundlot constraints is

$$\text{Minimize} \quad \frac{y^T Q y}{(e^T y)^2} \text{ s.t. } \bar{r}^T y \geq R_p, \ e^T y \leq 1 \text{ and } y_i \geq 0, \ i = 1, .., n \tag{44}$$

$$\text{and} \quad \phi(y_i) = 0, \quad \text{for} \quad i = 1, ..., n. \tag{45}$$

where $\phi(y_i)$ is given by (31). However this problem cannot be dealt with using DIRECT because of the unavoidable equality constraints (45). Since these constraints are also non-differentiable, the problem will not be easily solved by other standard constrained minimization methods. However we can set up another formulation using integer variables $z_1, ..., z_n$ which are related to the invested fractions, using the notation of section 4, via

$$y_i = z_i \frac{p_i}{M}.$$

The z_i must be non-negative and bounded above by the greatest integer which does not exceed M/p_i. Hence we can pose the problem as

$$\text{Minimize} \quad \frac{y^T Q y}{(e^T y)^2} \quad \text{s.t.} \;\; \bar{r}^T y \geq R_p, \;\; e^T y \leq 1 \tag{46}$$

$$\text{and} \quad z_i \;\text{ is integer } \;\; \text{and} \;\; 0 \leq z_i \leq \lfloor \frac{M}{p_i} \rfloor \;\; \text{for} \;\; i = 1, .., n. \tag{47}$$

This problem is in a form suitable for solution by DIRECT.

In conclusion, we mention some other methods for solving minimum-risk problems with buy-in thresholds and/or roundlot constraints. The problem defined by (6), (7) is a mixed integer quadratic programming problem (MIQP). The other problems in this section (i.e., (38), (39) and (40), (41) and (44), (45)) are general mixed integer nonlinear programming problems (MINLP). If the objective function is convex – as is the case for (6), (7) – then such problems can be solved by branch-and-bound techniques. These approaches involve solutions of a relaxed form of the problem *without* the integer constraint followed by solutions in which the z_i are fixed at integer values which bracket the non-integer solutions of the relaxed problem. If the objective function is non-convex then branch-and-bound can still be used, but convergence may only be to a local solution. A general purpose MINLP method is given by Fletcher & Leyffer (1994) while Mitchell & Borchers (1997) describe some branch-and-bound algorithms suitable for 0-1 MINLP such as (44), (45). Other software for solving MINLP problems is listed on http://www.gamsworld.org/minlp/solvers.htm.

Very efficient solution methods have been developed for mixed integer *linear* programming (MILP) problems and these can handle extremely large numbers of variables (Bixby et. al., 2000). Mitra et al (2003) quote results obtained with an effective MIQP solver called FortMP (Ellison et. al. 1999). While these involve problems which are not so large as those solved by MILP they are still considerably larger than the 50 variable example we have quoted in section 4 – and this remark raises a significant question about the specific approach that we have presented in the paper. It has never been claimed – even by its originators – that DIRECT is a suitable method for large-scale optimization. It is well-enough suited to problems in tens of variables; but, even though it lends itself quite well to parallel implementation, it does not seem likely to be developed into a method for problems in many hundreds of unknowns. Unfortunately, practical portfolio selection problems will typically involve hundreds, or even thousands, of assets.

DIRECT has played a valuable part in the work described in this paper because it has enabled us to do an initial feasibility study regarding the effectiveness of the nonlinear constraint function (8) for handling disjoint restrictions like (5). However, rather than applying DIRECT to the penalty function (11), it might be better to apply a general nonlinear programming algorithm (e.g. an SQP technique) to the problem (9), (10). Because this problem is likely to have many local solutions, it would be appropriate to solve (9), (10) within the framework of a multi-start approach, as proposed by Rinooy Kan & Timmer (1987a, 1987b). This method seeks the global optimum by performing many local minimizations from different starting guesses, using cluster analysis to ensure good coverage the region of interest by avoiding initial guesses that are too close together. On the basis of local minimizations already carried out, the method generates a Bayesian estimate of the number of minima that might still be undiscovered: and when this is sufficiently small the algorithm stops. Obviously there is a significant computational cost associated with finding many local solutions of (9), (10): but, equally, considerable effort may be expended in solving subproblems in a branch-and-bound approach to (6), (7). Hence, an interesting future project would be a study of methods for the minimum risk problem with buy-in threshold constraints: this would feature a numerical comparison between multi-start approaches involving continuous optimization problems like (9), (10) and branch-and-bound techniques for corresponding mixed-integer optimization problems such as (6), (7). Such an investigation could of course be extended to other portfolio selection problems with buy-in thresholds or roundlot constraints.

References

[1] M. Bartholomew-Biggs, S. Parkhurst and S. Wilson (2003), Global Optimization Approaches to an Aircraft Routing Problem, European Journal of Operational Research, 146, pp 417-431.

[2] M. Bartholomew-Biggs, Z.J. Ulanowski and S. Zakovic (2005), Using Global Optimization for a Microparticle Identification problem with Noisy Data, J. Global Optimization (32), pp 325-347.

[3] R. Bixby, M. Fenelon, Z. Gu, E. Rothberg and R. Wunderling (2000), MIP:Theory and Practice – Closing the Gap, In: M.J.D. Powell & S. Scholtes (eds), System Modelling and Optimization Methods Theory and Applications, Kluwer Academic Publications.

[4] C.G. Broyden (1970a), The Convergence of a Class of Double Rank Minimization Algorithms, Part 1, Journ. Inst. Maths. Applics **6**, pp 76-90, 1970

[5] C.G. Broyden (1970b), The Convergence of a Class of Double Rank Minimization Algorithms, Part 2, Journ. Inst. Maths. Applics **6**, pp 222-231, 1970.

[6] E. Ellison, M. Hadjan, R. Levkovitz, I. Maros, G. Mitra and D. Sayers (1999), FortMP Manual, OptiRisk Systems and Brunel University.

[7] R. Fletcher and S. Leyffer (1994), Solving mixed integer nonlinear programs by outer approximation, Mathematical Programming (66) p. 327

[8] N.J. Jobst, M.D. Horniman, C.A. Lucas and G. Mitra (2001), Computational Aspects of Alternative Portfolio Selection Models in the Presence of Discrete Asset Choice Constraints, Quantitative Finance **1**, pp 1-13, 2001.

[9] D.R.Jones, C.D. Perttunen and B.E. Stuckman (1993), Lipschitzian Optimization without the Lipschitz Constant, Journ. Opt. Theory & Applics, **79**, pp 157-181, 1993.

[10] D.R. Jones (2001), The DIRECT global optimization algorithm, In: C.A. Floudas & P.M. Pardolos, (eds), Encyclopaedia of Optimization, Kluwer Academic Publications pp 431-440.

[11] H.M. Markowitz (1952), Portfolio Selection, Journal of Finance, March 1952, pp 77-91.

[12] H.M. Markowitz (1959), Portfolio Selection: Efficient Diversification of Investments, John Wiley (1959), Second Edition, Blackwell (1991).

[13] J. Mitchell and B. Borchers (1997), A comparison of branch and bound and outer approximation methods for 0-1 MINLP, Computers and Operations Research (24) pp 699-701.

[14] J. Mitchell and S. Braun (2002), Rebalancing an Investment Portfolio in the Presence of Transaction Costs, http://www.rpi.edu/ mitchj/papers/transcosts.html, Rensselaer Polytechnic Institute, 2002.

[15] G. Mitra, T. Kriakis, C. Lucas and M. Pirbhai (2003), A Review of portfolio planning: Models and systems, In: S.E. Stachell and A. Scowcroft(eds), Advances in Portfolio Construction and Implementation, Butterworth-Heinemann, Oxford.

[16] A. Rinnooy Kan and G. T. Timmer (1987a), Stochastic global optimization methods. Part I: Clustering methods, Mathematical Programming (39) p.27.

[17] A. Rinnooy Kan and G. T. Timmer (1987b), Stochastic global optimization methods. Part II: Multilevel methods, Mathematical Programming (39), p. 57.