Portfolio Optimization without Programming or Lagrangians –Implications for Investing

Manuel Tarrazo
School of Management, University of San Francisco
2130 Fulton St.
San Francisco, CA 94117-1045
tarrazom@usfca.edu, (415) 422-2583 Office

ABSTRACT

We study alternative ways to optimize a portfolio to better assess the practical value of mean-variance optimizations. Mathematical programming methods do provide the correct solutions but in the process we lose sight of how security characteristics are reflected in the optimal weights. The first section of the study reviews pre-optimization challenges (data gathering, calculation of inputs). The second shows that early applications of mathematical programming to portfolio optimization were overloaded with features that are actually not needed in practice (e.g., semidefiniteness). This section also shows that through its history portfolio optimization has been hampered with inessential generality as well: the tendency to make things more "impressive" than they need to be to solve the problem. The third part of the study reviews alternative portfolio optimization methods –basis reduction, least squares, selective basis addition-- all of which can get started from an unrestricted quadratic equation, referred to as the algebraic method. The application of alternative optimization approaches to portfolio optimization provides twelve critical observations that reveal exactly how the optimization works. These findings enable investors not only to properly assess the practical value of mean-variance optimizations, but also to benefit from this knowledge without optimization calculations. Although the main focus is on enhancing investment practice and security selection, this study presents pedagogical advantages as well.

Key words: Portfolio optimization, mean-variance, Markowitz, quadratic programming, security selection.

INTRODUCTION

"Why don't Finance professors use portfolio theory?" --Doran and Wright (2010) ask in a recent study. They offer some interesting responses (some prefer to trade short term, others are passive investors, others have less investing experience than one might expect, and so on). In our opinion, not just Finance professors but also investors avoid portfolio theory because nobody knows what exactly portfolio optimizations do, or how optimal weights relate to security characteristics. Although we have at our disposal impressive-looking material that we can use to write imposing articles and to give weighty lectures, when it comes to investing precious household funds, of which hardly anybody has much to spare, reality checks in and mean-variance optimization is moved aside. This is the context in which this study was born: a practical need to enhance security selection by a household investor (small, individual).

This contribution clarifies mean-variance portfolio optimization to evaluate its application to actual, practical investing. Such an evaluation requires several steps --data gathering, computation of statistical inputs, optimization, and actual purchases of securities. Each of these steps can be evaluated by investors nearly instantaneously, with the exception of optimization, whose workings are still hidden in mystery.

In order to assess the practical value of portfolio theory we will be removing all the veils in portfolio optimization, an strategy that provides the best results for relatively small sized portfolios (i.e., limited diversification).

The optimization mystery is moving individual investors away from portfolio theory. Early contributions to portfolio theory, and especially that of Markowitz (1952), carried the promise of helping individual investors, a promise that started to die out almost equally from the beginning. Interestingly, some explanations of why this happened explain as well how optimizations became so hard to understand. One explanation is the intellectual success of portfolio theory, which generated so many contributions that the early interest in the individual investor was postponed. Another explanation is that, despite Markowitz's success at communicating the major principles, which his choice of a graphical approach made seem reasonably challenging, the investment problem was very difficult and it required highly specialized approaches, as in Roy (1952). And yet another is that the problem also requires including many other elements beyond household investors, for example markets and equilibrium considerations, as in Mossin (1966). It is interesting to observe that the move of portfolio theory away from serving individual investors coincides with the deepening of the mystery in the optimizations themselves. It is worrisome to think that a sometimes perceptible propensity in the literature for "the more complicated, the better" approach may have been at play, coexisting and often interfering with legitimate efforts to improve our understanding of the optimization. Some symptoms of such an obfuscating tendency are the fixation on more complicated methods when simpler ones will do, imposing concepts that actually impede practical applications, and the prevalence of pet-topics that can only exist in academia.

The first section of this study briefly reviews the pre-optimization challenges: a) the selection of the time-window for the sample, b) that of the number of securities, and c) the computation of statistical inputs. It is well-known that a) and b) are two major sources of doubt about the practical power of portfolio theory, for which the optimization itself has limited remedies.

The second section begins the analysis of challenges at the optimization stage and focuses on the methods employed at its early historical stage such as the critical line algorithm, the simplex method of quadratic programming, and Beale's procedure. The method of undetermined Lagrange multipliers takes the center stage next, after mathematical programming methods, especially for classroom purposes. The main focus of this second section, however, is to bring attention to those modifications to established methods that, nonetheless, were never taken advantage of despite their potential to simplify portfolio optimization, clarify what it really does, and help to evaluate the relationship between optimal weights and securities characteristics.

The third section presents three different ways to solve a portfolio without using programming methods or Lagrangians. This is accomplished by viewing optimization as ways to learn about the problem, and not merely as generators of computing solutions. We will show that a few insights, all of them used at one time or another in portfolio optimization, not only make traditional methods unnecessary but also fully reveal what the optimization does. More specifically, we will show that strict focus on what is practical (the no-short sales, mean-variance tangent portfolio, keeping utility implicit, and excluding risk-free rate and minimum variance issues initially), delivers the essence of what portfolio theory is: a value-preserving rule under a shared arbitrage procedure in a linear model. It also provides what investors need to know – captured in twelve critical observations -- to evaluate security trading using mean-variance methods. This critical know-how is further showcased in the fourth section of the study when contemplating actual investing periods. Concluding comments and references close the study.

The following references were useful at different stages of the work presented, although some of them may have not been explicitly quoted: Rubinstein (2006, 2002), who provides concise presentations emphasizing the intellectual impact of portfolio theory on Finance and financial economics in the field, see also Markowitz (1999) for a wider historical perspective; Roy (1952), Markowitz (1952, and 1959), and Tobin (1958) are considered the fundamental contributions to portfolio theory; recommendable modern review/surveys of portfolio analysis are those of Fabozzi, Gupta, and Markowitz (2002), Constantinides and Malliaris (1995), who focus on the investing component, and Steinbach (2001), who emphasizes the mathematical aspects.

Benninga (2008), reflecting theoretical, pedagogical, and applied concerns notes, “Markowitz changed the paradigm of investment management … nevertheless the MPT [modern portfolio theory] has disappointed … Anyone who has tried to implement portfolio optimization using market data knows that the dream is often a nightmare. Implementations of portfolio theory produce wildly unrealistic portfolios,” (op. cit. p. 349). This study shows that only when the veils are removed from the optimizations are we able to observe what they do, and to assess their practical value.

I. DATA AND OPTIMIZATION INPUTS

The first step in the evaluation of the practical value of portfolio theory is to collect the data. A common way to do so is to collect 5 years’ worth of monthly, price data for a number of stocks. This prescription is rarely found in the literature, with Fabozzi, Gupta, and Markowitz (2002, p. 9) representing an exception to the rule.

A few issues are finessed in one way or another at this stage, most notably a) how the stocks are selected, b) how many stocks are needed to get started, and c) data specifics, such as why monthly and why five years. At least in the classroom, it is very useful to address the first question by appealing to the analysis of fundamentals, which connects the exercise to courses on financial management and to one of the major approaches to investing (portfolio theory and technical analysis being the other two). Answers to the second and third questions usually touch upon statistics; thirty normally distributed variables, and even as few as ten on some occasions, allow for observing statistical regularities. We also use monthly data for regularity purposes, and we fear randomness as the frequency gets higher; perhaps months may allow for non-random effects to be felt with clearer distinctiveness. The five years issue relates to both statistics, that is having enough observations at the variable levels (60 observations), and also to capturing the current "structure" economic structure. There is a faint relationship with the economic cycle analysis, but portfolio theory is supposed to have a wider focus than the so-called market timing, see Fabozzi, Gupta, and Markowitz (2002) for further analysis.

The second step in the evaluation is to compute stock returns from the stock prices collected. The choice of discrete versus continuous formulations returns does not seem to pose major challenges. Including or excluding dividends is a more delicate matter. In the past it was very laborious to calculate dividends-adjusted prices and, perhaps because of that, the academic literature calculated returns based exclusively on closing prices. Nowadays, electronic reporting makes using dividend-adjusted closing prices, which are also adjusted for splits, easier.

The third step is to compute the statistical indicators proper –average returns for each 60-obs return series, and the variance-covariance (covariance heretofore) matrix for the m-securities selected, which represents the risk structure of the data. This step, which is made very easy with the tools provided by spreadsheet software, carries the information contained in the 60-by-m return matrix into a neat m-by-1 vector of returns, and an m-by-m risk matrix.

Before proceeding to the optimization stage, it is worthwhile to note that some of the most serious objections to using portfolio theory in actual security selection arise at this stage of collecting the data, and they will not be assuaged in any way by optimization. Rather, these objections will haunt the approach and cast doubt on its usefulness, seemingly permanently. At the heart of these concerns is the choice of a particular time-length for the data, which also implies a frequency choice. Markowitz (1959, p. 13 and ss.), for example, used 18 yearly return observations covering the period 1937-1954, which includes major social events (great depression, WW-II). Because of the statistical nature of the data, portfolio users implicitly accept that more is better than less in terms of both 1) observations, and 2) securities. The first preference brings to mind accuracy on parameter estimation which, with unchanging structures, is never hurt by having more observations. This, however, glosses over the extremely difficult issue of structural change and of capturing whatever is the most relevant structure for our investing purposes. Furthermore, "relevant" means "appropriate" in two senses: learning, in general, about the underlying eco-financial system, and also carrying out the optimization for the particular time-window of the investor. When the investor has a "time-window" to invest in equity (and everyone should have one according to common sense life-cycle reasoning) end-of-the-period wealth considerations matter more than sample variance, see Martellini and Urosevic (2006), Benjelloun (2010), Hsu and Wei (2003), and Jeffrey (1984) who emphasizes that the risk for small investors is simply the inability to make unavoidable payments.

The second preference –more securities are preferable to less-- refers directly back to the issue of diversification, which is often handled in statistical terms, and nothing is said about the quality of the securities, see, for example, among many articles the following: Domian, Louton, and Racine (2007), Boscaljon, Filbeck, and Ho (2005), Shawky and Smith (2005), Newbould and Poon (1993), Statman (1987), all the way to the original study by Evans and Archer (1968) and it modern re-evaluation by Bellenjoun's (2010).

Evans and Archer (1968) found that even a ten-security portfolio of randomly selected securities could offer as much risk reduction as non-randomly selected portfolios of many more securities. The statistical angle, however, may have little to do with creating a successful portfolio because returns can be eaten away by over-diversifying ("diworsifying" according to Peter Lynch's expression). Moreover, what investor is going to place a dear amount of funds in randomly selected stocks? Mentioning individual investors raises yet other concerns to the unfinished business of data length, its frequency, and the number of initial stocks to prepare for optimization. These additional concerns have to do with individual investor characteristics (age, human and non-human wealth, and investing for a given goal or purpose, which are beyond the focus of this study –see, however, Martellini and Urosevic (2006), Lee (1990), Jaganathan and Kocherlakota (1996), Lee and Hanna (1995), and Coger and Ruland (1983).

In sum, before passing on to the optimization it is appropriate to close this section by keeping in mind the following reflections:

a) "There is certain arbitrariness to measuring returns simply as a function of units of time. In some periods, no significant events will take place to cause prices to change, so returns will essentially reflect noise. In other periods, several important events will influence returns. But the typical estimation of risk parameters assigns as much importance to the periods with no significant events as it does to the event-filled periods." Chow et alia (1999, p. 65).

b) "The truth is that there is no right answer because we are dealing with the world of uncertainty," Fabozzi, Gupta, and Markowitz (2002, p. 10).

II. MEAN-VARIANCE OPTIMIZATION –STANDARD APPROACHES

Let C be the m-by-m covariance matrix, r the m-by-1 vector of average returns, and x_i represent the portfolio weight for security "i". Markowitz's mean-variance formulation can be expressed in the formula group (1):

$$\text{Min} \quad x'\,C\,x \tag{1a}$$
$$\text{St} \quad x'\,r \geq rp^{*} \tag{1b}$$
$$w'\,\mathbf{1} = 1 \tag{1c}$$
$$x_i \geq 0 \tag{1d}$$

Where σ^2_p = x' C x is the portfolio variance; rp = x' r is the portfolio return; rp* represents the numerical value for required return set by the user; and w' **1** = Σ_i w_i = 1 is the "full investment" assumption.

This specification seems very similar to standard specifications for quadratic programming (QP) in formula group (2), as well as Wolfe's specification in formula group (3). There are, however, subtle differences that make a real difference when trying to understand what the optimization does.

$$\text{Min} \quad F(x) = \tfrac{1}{2} x' Q x - x' p \tag{2a}$$
$$\text{s.t.} \quad A^T x \leq b \qquad \text{(inequality constraint)} \tag{2b}$$
$$E^T x = d \qquad \text{(equality constraint)} \tag{2c}$$
$$x_i \geq 0, \text{ all } i \tag{2d}$$

$$\text{Min} \quad F(\lambda, x) = \lambda\, p\, x + \tfrac{1}{2} x' C X \tag{3a}$$
$$\text{S. t.} \quad A x = b \tag{3b}$$
$$x \geq 0 \tag{3c}$$

Beale (1955), further detailed in Beale (1959), seems to have been the earliest contribution providing solutions to minimize a convex function subject to linear inequalities. Markowitz (1956) proposed the critical line algorithm that seems custom-made for the portfolio optimization problem. Wolfe (1959) delivered the algorithm that became most popular (simplex method for quadratic programming). Comparative analyses of each method can be found in Simmons (1975) and in the more recent survey of the field by Nocedal and Wright (1999).

As we noted earlier, Markowitz (1952) successfully used graphical explanations of portfolio selection, and tried to use the same strategy when describing the critical line method: "The required principles of vectors and matrices are not difficult to learn, even for the reader with a meager mathematical background," Markowitz (1959, p. 154). Despite friendly invitations, the difficulty of the problem, and the inherent complexity of mathematical programming, could not be ignored. The state-of-the-art provided by these contributions proved prohibitive for most researchers and practitioners. Problems with minimal dimensions (two or three variables) require pages and pages of explanations, even in a scholarly medium that is known for its conciseness. Further, even in the case of the more forgiving contribution, Wolfe (1959), the user ends up optimizing a set of variables twice removed from the original ones, which makes the investor lose sight of how the properties of the original variables play out in the optimization. Even to this day, these early programming methods are hard to understand by non-specialists. The critical line method, for example, seems to lead its user through the optimization and lends itself well to graphical illustrations, but still retains its difficulty despite excellent explanatory treatments such as that of Kwan and Yuan (1993).

Here we come to a problem that has plagued portfolio optimization until very recently: the traditional computation procedure requires very specialized knowledge and it also hides why different securities have different weights or, in other words, how one asset is preferred to another. This unfortunate outcome results from at least two factors; One of them is the excess baggage originated by mathematical programming, which is set up, a priori, to handle many potential cases that may not materialize in the optimization (semidefinite functions, non-negativities, inequality constraints, restricted variables that need not be restricted, and so on).

The second contributing factor is the observable and unfortunate practice in the literature of favoring the complicated over the straightforward. In the end, authors after authors have equipped the optimization to cover each conceivable potential case, no matter how unlikely. This excess baggage and "inessential generality" buries financial logic in numerical manipulations that become incomprehensible to the investor.

This is a brief listing of the excess baggage mentioned above:

a) **Range and definiteness of the function.** In some cases, the excess baggage is triggered by a very reasonable concern. Take, for example, this observation by Markowitz: "Suppose further that some decision-maker likes expected pay-off (E) and dislikes variance of pay-off (V). Our problem is to compute for the decision-maker … The 'efficient combinations' of E and V, i.e., those attainable (E, V) combinations which give minimum V for given E …," Markowitz (1956, p. 111). Addressing this question requires one to consider zero values for some range in the solution space for the functions involved; this is the issue of semi-definiteness versus definiteness. In the case of portfolio optimization, the quadratic form x' C x in (1a) is positive definite, which means that any non-zero (non-trivial) value for x, the vector representing optimal weights, will yield a positive variance value. This is a fact which, if properly exploited, simplifies the optimization drastically, as studied by Wolfe (1959) who provided both "short form" and "long form" algorithms for the positive definite and the semidefinite case, respectively.
   Note further that the reasonable question above introduces an inequality constraint for the required return of the portfolio (1b). This is not necessary because the optimization will happen with equality, and it has two effects. One is that it brings another variable into the optimization. The second effect is that, even setting this constraint as an equality, it still introduces definiteness considerations. Even when Q is positive definite, the matrix M = $[Q\ A^T;\ A\ 0]$ employed in the computation of solutions is not even semidefinite –see, for example, Luenberger (1984, p. 431), which may bring about numerical difficulties. (By the way, in many cases, maximizing is usually easier than minimizing. Further, in the case of portfolio optimizations, there is some intuitive value in maximizing returns and maximizing the negative of the variance.)

b) **Neglect of the unrestricted formulation tying risk and return**. If we compare Markowitz's and the general quadratic programming (QP) specification, we notice that the objective function in the QP is a quadratic equation, composed of its quadratic form and its associated first order vector. As noted earlier, in the context of portfolio optimization, these correspond to the portfolio variance and return, respectively. Moreover, they acquire special meaning when they are together because they represent "market conditions", without the interference of any particular investor's imposed conditions. As noted earlier, Markowitz, although he was writing at the birth of modern conceptualizations of "market conditions", focused on the individual investor.

Interestingly, Wolfe's specification included a lambda (λ) term, which indicates parametric programming methods are being used. By coincidence, the same parameter can represent a risk-aversion coefficient ($\lambda = 0$, maximum concern about risk; $\lambda = 1$, full concern with returns) that can replace the required return restriction. This coincidence has the effect of bouncing the analysis back to the whole potential range rather than to the unrestricted problem represented by the quadratic equation.

c) **The sum of weights equal to one constraint** can also be omitted if the weights are simply rebalanced to add up to one in the restricted equation, as we shall show in the second section.

d) **The non-negativity requirement** concerning choice variables can also be applied on an "as needed" basis after running the unrestricted optimization.

It turns out that some writers did indeed try to find alternatives to the full-fledged programming approach. Martin (1955) suggested a procedure to deal with negatively-signed weights by eliminating from the basis (and programming tableau) those problematic variables. He suggested this at seemingly the best time to do so, when interest in portfolio theory was exploding. His method had excellent credentials building from Markowitz's critical line algorithm and appeared in a well-respected journal. Only Francis and Archer (1971), with their characteristic thoroughness and practical mindedness, revisited Martin's method, as we shall note later. Martin's (1955) neglect seems costly today: without non-negativity constraints, researches would perhaps have found it obvious that the equality constraint ($\Sigma_i\ w_i = 1$) did not change the problem structure by itself, and amounted to a recalculation of weights that could be performed after running the unrestricted optimization. Lemke (1962) concentrated on exploiting the quadratic equation and the unrestricted case first, and then added restrictions as needed in a particular sequence. Later on, Beale reiterated the need to lighten up the optimization: "It is, therefore, natural to wonder whether, instead of carrying around all this information in case it is needed, one cannot represent the problem more compactly, calculating particular elements of the tableau only when required. It turns out that this is possible," Beale (1967, p. 164). Unfortunately, the efforts of these authors trying to increase our knowledge of the problem by clarifying the optimization itself were not taken advantage of. It is customary even nowadays for authors to begin their portfolio optimization research invariably with the full-fledge regalia of the early programming methods, despite appeals for clarity: "The mathematics are so complicated and/or so abstract as to not to be easily understood or responded to in an intelligent way by the persons who are responsible for making the crucial value judgments," Renshaw (1967). Portfolio theory had an "extremely unwieldy nature," for Dickinson (1974, p. 447).

The situation was understandable for some time, because portfolio theory was a privileged child of what has been described as a "golden age" for Economics, as evidenced in Samuelson (1983, originally published in 1947). Game theory, input-output analysis and mathematical programming were revolutionizing our understanding of problems that, like portfolio selection, were theoretically difficult and had practical applications. The developments with respect to tools were also very impressive. The gap between calculus (analysis, continuous functions) and operational methods (the algebra of inequalities, matrix and numerical procedures) was reduced to a well-built and travelled bridge where techniques were provided for obtaining global solutions under a fairly wide range of conditions covering most cases of theoretical and especially practical interest, see Pardalos and Rosen (1987) for an analysis of the techniques that flourished with portfolio theory.

The situation, that is, overly equipping the portfolio optimization, was also understandable when considering the challenges presented by the optimization of a portfolio. Markowitz (1959, p. 383) reminds that the state-of-the-art of astronomical computations were hardly capable of handling a 25 securities problem at the time. Still, 50 years later, Steinbach's decision (2001, p. 34) is as useful as it is uncommon: "We give higher priority to a clear presentation, and inessential generality will sometimes be sacrificed for technical simplicity. In particular, no inequality constraints are included except when necessary."

In 1960, right at the apogee of mathematical programming research, Theil and Van de Panne (1960) published research that would establish the way portfolio optimization is taught in textbooks to this day. Their contribution was one of those bridging classical (calculus) and programming methods. They showed that a conventional quadratic maximization applying the method of Lagrange would yield the same optimal solutions as quadratic programming. The following formula is a typical characterization of the method of Lagrange as it appears in investment textbooks --see for example Luenberger (1998, pp. 158-160):

$$L = \tfrac{1}{2}\, x'\, C\, x - \lambda_1 (x'\, r - rp^*) - \lambda_2 (x'\, \mathbf{1} - 1) \tag{4}$$

This corresponds to

Minimize	Portfolio variance = varp = x' C x
Subject to	Required return; x' r = rp*
	Sum of optimal weights equal to one; x' **1** - 1

The first order conditions for a 3-stocks case can be written as

$$\begin{vmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & -r1 & -1 \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & -r2 & -1 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & -r3 & -1 \\ r_1 & r_2 & r_3 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda_1 \\ \lambda_2 \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ 0 \\ rp^* \\ 1 \end{vmatrix} \tag{5}$$

Note the distinctive features of this optimization. First, a quadratic form, x' C x. Second, the so-called Lagrange "multipliers ($\lambda_1$and λ_2), which replace each equality constraint for a variable in the objective function of the optimization, equation (4). These multipliers represent sensitivity factors (first derivatives of the objective function with respect to the specific constraint). Therefore, they function like brakes (when maximizing), or like floaters (when minimizing); the higher their value the more they restrict the function to be optimized. Third, in the case of portfolio optimization, sufficient conditions for the extreme are met because of the positive definiteness of the covariance matrix C, which is always strictly positive for any non-trivial vector of optimal weights. Fourth and final, the optimal solutions are obtained with a simple simultaneous equations system.

The advantages offered by classical optimization over mathematical programming formulations are critical, especially in the classroom: The problem is easier to understand, and only two auxiliary variables enter in the picture, no matter how many stocks are optimized. As noted earlier, most students are familiar with and have solved simultaneous equations system, which also have valuable numerical properties in the case of portfolio optimization (e.g., convergence). In addition, it turns out that Lagrange multipliers are very familiar to portfolio optimization users because they had become important in economics and business analysis –e.g., identical optimization methods become pervasive in microeconomics, where they may play a role of "shadow-prices" championed by Samuelson (1983).

Unfortunately, this method has also been subject to a) overburdening, b) misuse, and c) and neglect, as was the case with some of the developments in quadratic programming (Martin, Lemke, and Beale). The overburdening comes again from injecting utility concepts into the optimization. For example, the Lagrangian function can be written as

$$L = [(1- \lambda_1) (½\, x'\, C\, x)] + (\lambda_1)\, x'\, r + \lambda_2 (x'\, \mathbf{1} - 1) \tag{6}$$

This specification is attractive because the first Lagrangian takes the role of a coefficient of risk sensitivity or aversion; $1 \geq \lambda_1 \geq 0$, when $\lambda_1 = 0$ ($\lambda_1 = 1$) variance (return) is the only concern. Moreover, this specification brings to mind Wolfe's parametric programming approach to quadratic programming. The wider generality, however, veils the role of the Lagrangian multipliers as well as how the solutions relate to security characteristics. It may also be misleading to assume the investor has a choice concerning risk and return. Finally, as we will note in the next section, wider generality may also hinder understating implied arbitrage relationships.

The classical-Lagrange specification in (4) is also frequently abused. For example, as has been noted by Steinbach (2001, p. 39), when this model is used with two variables and two restrictions it permits no choice. In this case, the model provides simply an average of returns. Worse yet, the solution will exclude the risk matrix. See Tarrazo and Murray (2004) for a criticism of the two-variable, two-restriction model in the analysis of asset allocation, where the variables represent stocks and bonds. Very often, perfect correlation is assumed in illustrations of a two-variable model, but this does not make any sense because it would make the matrix C singular. One can also find cases where it is assumed a non-stochastic variable in C, and even more misuse piles up when arbitrary changes are made in the return vector or in the risk matrix, forgetting about their simultaneity.

Another instance of analytical abuse happens when the simplifying tool is used to complicate matters. For example, it can be shown that in the minimization of $F(x) = ½\, x'Q\, x + cx$, subject to $A\, x = b$, the solution can be expressed as

$$x = - Q^{-1} [I - AT\, (AQ^{-1}AT)^{-1}\, AQ^{-1}]\, c + Q^{-1}\, AT\, (AQ^{-1}\, AT^{)-1}\, b \tag{7}$$

While this road may be helpful in resolving other problems, it is not so in portfolio optimization because the path to finding solutions is unlikely to show how they relate to security characteristics.

Finally, as was the case in mathematical programming (Martin, Lemke, Beale), the Lagrangian framework is yet to be fully exploited. The neglected work in this case is that of Boot (1964, 1963, and 1962), some of which has the potential to drastically simplify and clarify the classical-Lagrangian setup even further, as we will show in the next section.

In sum, Lagrange methods became justifiably the textbook standard, but their promise to simplify and clarify the relationship between solutions and security characteristics did not fully materialize; somewhere in the optimization, the rationale for preferring one security over another remains hidden. And the issue of non-negativity still needed to be addressed. In the next section, we will show how alternative optimization methods will take care of this problem. Further, we will also solve the major problem of portfolio optimization: observing exactly how different securities have different weights and how one asset is to be preferred over another. In order to learn that we only need to simplify and try to learn from alternative optimization settings.

III. OPTIMIZATION ALTERNATIVES AND WHAT THEY SHOW

In the previous section, we reviewed the development of optimization methods for mean-variance portfolios. What appeared initially to be a rather difficult problem found a manageable solution in Wolfe's (1959) method, which seemed reasonable in its demands on the user, at least when compared to the alternatives, and quickly took the center stage in financial practice. Similarly, the classical optimization-based method of Lagrange took over classroom illustrations of portfolio optimization. Nevertheless, neither of these procedures has helped to understand what the optimization does, or why a security is preferred to another. We ventured several reasons for this: a) some interesting leads were either simply ignored or not followed up; b) the optimization was overwhelmed with too many unnecessary elements that veiled its workings;; and, c) nobody tried to solve the problem by a better understanding of its economic and financial content, all the effort being put into the numerical aspects of the optimization.

In this section we show that, while the mean-variance portfolio optimization is indeed a quadratic optimization problem, there is no need –at all— to use quadratic programming or classical- Lagrangian methods to solve it. Moreover, finding the optimal solution by alternative methods not only reveals what the optimization does, but also provides valuable practical investing implications.

Our first step is to view optimization as a way to learn about the problem to be solved. A well-known author in the optimization literature seems to have recognized our same need when he entitled one of his studies, "(T)he purpose of mathematical programming is insight, not numbers, " Geoffrion's (1976). He noted further, "they [mathematical programming methods] do not explain WHY the solution is what it is," (1976, p. 81). The objective, then, is to develop insight that will ultimately be applicable to practical decision-making, especially when problems remain after optimal solutions have been calculated.

There are ways to develop insights from optimization efforts that have proven successful in the literature: 1) learn from different mathematical programming methods to see if and when these methods can be avoided; 2) look for in other optimization approaches (taboo searches, fixing and excluding variables, visual representations, graph theory insights, and so on); 3) consider preprocessing –e.g., rearranging the data; and, 4) deploy heuristics such as problem decomposition, sequencing, data partitioning, inductive methods; and, in cases where exact solutions are not existent or too expensive to obtain, 5) consider approximations, local improvements, and constructive techniques that outline likely solution spaces, see Ball (2011), Silver (2004), and Nemhauser and Wolsey (1988). We will proceed to do so.

III.1 Algebraic Formulation Of Portfolio Optimization

Our first strategic move is to focus on what we will refer to as the algebraic formulation of portfolio optimization, which is a quadratic equation composed by a quadratic form (x' A x) and a vector (x' b). Of course, A and b take the place of the covariance matrix and average individual stock returns, respectively, but the analysis is brought to the well-known area of simultaneous equations systems (SES).

$$F(x) = - \tfrac{1}{2}\, x' A x + x' b \tag{7}$$

Next, we remove any reference to a desired/required return or to utility theory. The fundamental tenets of portfolio risk-return choice still remain in that equation, but this time implicitly, as done in Constantinides and Malliaris (1995, p. 4). We also solve for the best available portfolio given the data provided and we will arrive to the no-short sales (NSS, $w_i \geq 0$ for all securities) best portfolio solution. By doing so, we avoid the dust-storms raised by utility theory, we keep the analysis direct, honest and humble (one needs to know a lot to be able to impose a required return on a portfolio), and we are able to work with market data. Note also that we begin by not including the "sum of weights equal to one" restriction beforehand.

Exhibit I presents the classical-Lagrangian optimization, which is tried out with an arbitrary 0.02 required monthly return (2%, 24% annual). The northwest part shows the results of several optimizations. Observe the values for the Lagrangian multipliers appearing just below the optimal weights. For the 2% required return both of them are different from zero. There are, however, two distinct cases where one of them is exactly zero –one of them corresponds to the minimum variance portfolio (minvar), the other to the tangent portfolio, which is the one that maximizes the return-to-risk ratio of the portfolio (rp/pvar). The returns for those cases can be found by a trial and error method (e.g., "goal seek" of EXCEL). The tangent portfolio permits us to pass from the restricted, classical-Lagrangian model in (4) to the unrestricted, algebraic specification in (7). In effect, let's take a look around the same area in Exhibit II. We will see the same optimal weights, which we calculate from the auxiliary variables solving A x = b simply by rebalancing, $w_i = x_i/\text{sum}(x)$. Note that sum(x) = rp/pvar = $1/\lambda1$ = 6.832443. This number is the most important in the optimization. It appears as the bottom of the “cup” represented by the quadratic form in (7) –that is, it is also the return-to risk ration of the minimum variance portfolio. Normally, it would appear with a negative sign, which would identify it as the (conjugate) gradient: the direction that would lead into the global optimum starting from the bottom of the cup, given the forcing vector b; in other words, the optimal “ascent” ratio. (The negative sign would come from maximizing the negative of a “bad” thing --the risk term-- for which we would write - ½ x’ C x. In this study, however, we adopted Luenberger’s formulation, equation (7), of the classical-Lagrangian, for the reader’s convenience. In addition, we show non-percentage numbers so that the reader can appreciate numerical accuracy.)

Before eliminating securities with negative weights, let us make our first two critical observations:

Observation 1: Equality of marginal (security level) and total conditions (portfolio level)

The SES approach conveniently illustrates the relationships between total and marginal conditions. Let us see: we have found the solution for x, such that A x* = b, for example by matrix inversion –A is invertible, determinant not equal to zero, and A and its inverse are also positive definite. This vector of solutions x* is real-valued, which means that each number is positive, negative, or exactly zero. For numerical reasons, it is difficult to observe securities with exact zero weights. If found, they are “said to be out of the portfolio,” Sharpe (1991, p. 501), and can be eliminated from the system. We would have A x* = b, and –because of linearity-- the following is true as well: x’ A x = x’ b. This amounts to multiplying each original equation for x_i and adding them up. After the multiplication, each equation shows the marginal variance of the security (as represented by its auxiliary or “subrogate” variable) being equal to the marginal return, which is the average return of each security. Note also, because of linearity, the ratio of the marginal return to marginal variance is the same for each of the optimal portfolio weights as well: $r_i/C_iw_i^* = 6.83244 = 1/\lambda 1$, as shown to the right of the optimal weights in the Exhibit II. C_i is the i_{th} row of the variance covariance matrix, which transposed and pre-multiplied by the optimal weight vector provides the marginal variance for the “i_{th}” security.

For any non-trivial vector of portfolio weights the variance of the portfolio will be positive. Add to this fact that a positive portfolio return will have a positive return-to-risk ratio that must also be met by each individual security. If the portfolio (total, portfolio level) return-to-risk ratio must be equal to that of each security (marginal, security level), and if the portfolio has a positive ratio, wouldn’t we expect securities with negative returns to be excluded from long-only (no short-sales) portfolios? As we will see shortly, this is shown to be the case when we apply a basis reduction method to the SES solution.

Observation 2: The tangent portfolio a) is the unique optimal global solution, b) has the highest risk-to-return ratio, and c) is the only one endowed with market arbitrage conditions.

Exhibit I. Classical-Lagrange Portfolio Optimization and Basis Reduction.

	MCD	IBM	HPQ	KO	PFE	GE		ri	std.P	ri/stdi
MCD	0.002337	0.000917	0.001509	0.001356	0.001494	0.002187		0.016911	0.048345	0.349792
IBM	0.000917	0.003591	0.002336	0.001217	0.000315	0.002359		0.009801	0.059928	0.163552
HPQ	0.001509	0.002336	0.004872	0.001488	0.001697	0.003797		0.010851	0.069798	0.155465
KO	0.001356	0.001217	0.001488	0.00241	0.001117	0.002856		0.005473	0.049096	0.111483
PFE	0.001494	0.000315	0.001697	0.001117	0.004184	0.003008		-0.00709	0.064684	-0.109601
GE	0.002187	0.002359	0.003797	0.002856	0.003008	0.010211		-0.01148	0.101047	-0.113582

	MCD	IBM	HPQ	KO	PFE	GE				
MCD	0.002337	0.000917	0.001509	0.001356	0.001494	0.002187	-0.01691	-1		0
IBM	0.000917	0.003591	0.002336	0.001217	0.000315	0.002359	-0.0098	-1		0
HPQ	0.001509	0.002336	0.004872	0.001488	0.001697	0.003797	-0.01085	-1		0
KO	0.001356	0.001217	0.001488	0.00241	0.001117	0.002856	-0.00547	-1		0
PFE	0.001494	0.000315	0.001697	0.001117	0.004184	0.003008	0.007089	-1		0
GE	0.002187	0.002359	0.003797	0.002856	0.003008	0.010211	0.011477	-1		0
	0.016911	0.009801	0.010851	0.005473	-0.00709	-0.01148	0	0		0.02
	1	1	1	1	1	1	0	0		1

	SS allowed				NSS
Tickers	Initial	minvar	tangent	BR1	BR2, tngnt
MCD	0.703426	0.281704	1.644732	0.910016	0.874922
IBM	0.232283	0.281711	0.121958	0.166776	0.125078
HPQ	0.110642	-0.00287	0.364005	-0.07679	
KO	0.251291	0.364053	-0.0004		
PFE	-0.03742	0.231704	-0.63813		
GE	-0.26022	-0.1563	-0.49216		
λ1	0.045284	-1.3E-08	0.14636	0.127968	0.127708
λ2	0.000974	0.001411	2.57E-09	5.9E-17	-4.49E-10
	1	1	1	1	1
(Monthly)					
rp	0.02	0.009638	0.043129	0.01619	0.016021
pvar	0.00188	0.001411	0.006312	0.002072	0.002046
pstd	0.043357	0.037558	0.079451	0.045517	0.045233
rp/pstd	0.461287	0.25661	0.542843	0.355695	0.354195
rp/pvar	10.63929	6.832431	6.832443	7.814468	7.830387
(Yearly)					
rp	24.00%	11.57%	51.76%	19.43%	19.23%
pvar	0.022558	0.016927	0.075749	0.024862	0.024553
pstd	15.02%	13.01%	27.52%	15.77%	15.67%
rp/pstd	1.597945	0.888923	1.880463	1.232164	1.226968
rp/pvar	10.63929	6.832431	6.832443	7.814468	7.830387
mdeterm	8.19E-14	8.19E-14	8.19E-14	1.69E-07	5.05E-05

Exhibit II. Algebraic Model Optimization and Basis Reduction.

	MCD	*IBM*	*HPQ*	*KO*	*PFE*	*GE*
MCD	0.002337238	0.000917	0.001509	0.001356	0.0014941	0.002187
IBM	0.000917225	0.003591	0.002336	0.001217	0.0003145	0.002359
HPQ	0.001508886	0.002336	0.004872	0.001488	0.0016966	0.003797
KO	0.001356104	0.001217	0.001488	0.00241	0.0011171	0.002856
PFE	0.001494109	0.000315	0.001697	0.001117	0.004184	0.003008
GE	0.002186659	0.002359	0.003797	0.002856	0.0030082	0.010211

	xi	wi		Ci = C w	ri/Ci = 1/λ1	Ci/ri = λ1
MCD	11.23754122	1.644735		0.002475	6.8324329	0.146361
IBM	0.833270815	0.121958		0.001435	6.8324329	0.146361
HPQ	2.487044177	0.364006		0.001588	6.8324329	0.146361
KO	-0.0027303	-0.0004		0.000801	6.8324329	0.146361
PFE	-4.36001433	-0.63813		-0.00104	6.8324329	0.146361
GE	-3.36267871	-0.49216		-0.00168	6.8324329	0.146361
	6.832432876	1				
mdeterm	5.04895E-16					

	MCD	*IBM*	*HPQ*			Returns
MCD	0.002337238	0.000917	0.001509		MCD	0.016911
IBM	0.000917225	0.003591	0.002336		IBM	0.009801
HPQ	0.001508886	0.002336	0.004872		HPQ	0.010851

	xi	wi
MCD	7.11128813	0.910016
IBM	1.303264617	0.166776
HPQ	-0.60008478	-0.07679
	7.814467965	1
mdeterm	2.23291E-08	

	MCD	*IBM*				Returns
MCD	0.002337238	0.000917			MCD	0.016911
IBM	0.000917225	0.003591			IBM	0.009801

	xi	wi
MCD	6.850974474	0.874922
IBM	0.979413013	0.125078
	7.830387487	1
mdeterm	7.55255E-06	

The proof of a) is one of those neglected leads in portfolio optimization literature. In this case the lead is due to Boot. Take a look at the Lagrangian optimization in Exhibit I. As noted earlier, there is a certain value for the required return set by the investor of (rp* = 0.043129) that makes the second Lagrangian ($\lambda2$, the one implementing the full investment restriction) equal to zero, i.e., unnecessary. Boot's research (1964, 1963, and 1962) encountered this situation which, at first, appears to be an "anomaly". Initially, he defined this type of constraint: "A constraint is trivial if, and only if, whenever the remaining constraints are satisfied, the constraint itself is also always satisfied," Boot (1963, p. 420). Then, he proved that the correct procedure to be applied to a constraint is to check it for triviality and eliminate it from the optimization … "before continuing to check the other constraints for triviality (!)," (ibid, p. 421). And finally, Boot found that the very presence of that constraint being trivial and with the right sign (negative), actually confirms the uniqueness and global optimality of the solution and, therefore, implements duality theory (Farkas theorem, Kuhn-Tucker conditions) as well.

In our case, that the constraint becomes trivial at the critical rp* also proves that arbitrage is unrestrictedly at play. The first restriction is fully effective, which is the restriction embodying arbitrage: "This corollary [stating that a portfolio's return equals the weighted returns on the assets included in the portfolio] is often used in the literature without the realization that it is actually a consequence of an economy with no arbitrage opportunities," Jarrow (1988, p. 24).

Arbitrage is the economic force integrating securities and portfolios, and it acts freely in the tangent portfolio, which embodies the no-arbitrage conditions and is also the solution to the unrestricted quadratic equation (7). This equation contains only unaltered market data, which means the arbitrage acting in the optimization is market arbitrage. There is a risk that imposing ill-informed restrictions, such as an arbitrary required rate of return, may affect the informational content of the market data.

Additional detail concerning arbitrage can be found at Ellerman (1990, 1984), who studies the arbitrage implications of classical optimization and the role played by Lagrangian multipliers and the mathematical treatment of arbitrage, respectively.

Again, the no-short sales (NSS) portfolio with the highest rp/pvar ratio is called the tangent portfolio. We base portfolio theory on the premise that "investors will prefer more returns to less, and less risk to more." Therefore, it follows that the (NSS) tangent portfolio should be best for any type of mean-variance investor because the other portfolios are worse, period. This portfolio also carries no-arbitrage market information, and it makes the least informational demands on the investor. However, for some reason, the literature bends over backwards to justify other portfolios.

Up to this point, we have shown that the algebraic method can replicate the results of the classical Lagrangian without using any restrictions. We will address the issue of non-negative restrictions concerning optimal portfolio weights next.

III.2 Basis Reduction

It turns out that all we need to do to get the no-short sales (NSS) solution is to eliminate troublesome securities from the basis, like peeling an onion. This amounts to changing the dimension of the basis, which we can do because the function is quadratic. Exhibits I and II show this “Basis Reduction” (BR) method at work. This is something that both Martin (1955) and Beale touched upon. Martin’s (1955) strategy suggestion was not pursued by other researchers, which may have had been because it was formulated in the context of the critical line method, where the contribution’s suitability may have been hard to appreciate. One exception was Francis and Archer (1971): “It is possible to extend this algorithm [Lagrangian method] so it does not produce negative weights. At the point where the first weight reaches zero (before becoming negative) stop the analysis. Remove the row and column...” Francis and Archer (1971, p. 84). Another exception is Tarrazo (2000), who provides an analysis of Martin’s basis reduction procedure and shows how it contributes to our learning about portfolio optimization.

The basis reduction method seems an obvious choice to use in the cases of the classical-Lagrangian or algebraic methods, as shown in Exhibit II, as it unveils a great deal about how the portfolio optimization works.

> **Observation 3**: securities with negative weights in the unrestricted optimization can be removed from the basis.

Tarrazo (2000) explains in different ways why this is (and should be) so. Some of the explanations come from linear algebra: the signs are well-defined for real numbers; values for the negative variables can be set to zero; a column of zeros can be omitted from the matrices. Others, come from programming theory –e.g., identify sign-unrestricted variables and run an unrestricted optimization for those variables, Luenberger (1973, especially chapter 5), and Tarrazo (2000, p. 46).

> **Observation 4**: Securities with negative returns will not enter in the optimal portfolio if this portfolio has positive returns.

We saw earlier that linearity requires equality between the return-to-risk ratio for the portfolio and the marginal return-to marginal risk ratios for the securities. At the portfolio level, any vector of non-zero weights generates a positive portfolio variance; it does not make sense to assume an investor would bear this risk without a positive return. Therefore, each component (numerator and denominator in rp/pvar) is positive. At the level of individual securities, it is logical to expect that pressing for a higher return would cause a corresponding increase of risk. Therefore, for an optimal security, one would expect a positive (and increasing) marginal variance attached to a positive marginal return, which is simply the average return of the security. This logical explanation supports observation (4) above. But what about the covariances? Tarrazo (2008) uses the concept of quasi-diagonal dominance to prove they will not interfere with the logical argument above.

A matrix is diagonally dominant when the sum of absolute values of off-diagonal elements (covariance) is less than the diagonal element (variance). A matrix is quasi-diagonal dominant when there is a vector of weights (x, w) that makes the matrix diagonally dominant. Covariance matrices are positive definite, which means they have a "weighty" diagonal that makes the determinant always positive, but they are not diagonal dominant. This, in turn, means that the solution vector in (7) plays two roles, one ensuring that the total and marginal conditions are maintained both at the return, and at the risk conditions by making sure that marginal variances are positive as well, for optimal stocks. The logical argument stands. Difficulties, however, do not arise from the risk-side but from the returns-side.

> **Observation 5**: Positive security returns are a necessary, but not a sufficient, condition for a positive portfolio weight.

It is well-known that when negative weights (short-selling) are allowed, stocks sold short bring financing to buy the stocks with positive returns. Drying the financing provided by stocks sold short shows which stocks benefit from it and which would otherwise be not included in the optimal no-short sales portfolio. In sum, despite their positive average return, some securities are simply not worth holding without the money brought in by short selling.

Tarrazo (2000) shows that drying the financing provided by the shorts can easily be done by adding linear restrictions requiring that the weights for securities other than the short add up to one. Adding these auxiliary restrictions provides the no-short sales optimal solution. But instead of adding restrictions, it is easier to eliminate those troublesome securities, as is done in the basis reduction method. This means portfolio optimizations can be dramatically improved by initially discarding securities with negative returns. If the reader looks very closely, some authors employ only positive return securities in their analysis, starting with Markowitz (1959, Chapter 2) or, alternatively, all positive weights Tobin (1958, p. 82), which clearly has the same implications. Interestingly, Markowitz (op. cit.) uses yearly data, which makes it easier to find positive average returns.

It seems the fact that negative returns will never carry positive weights is a hidden secret only known to true initiates. Why is that so? Is it because this fact shows a fatal weakness in mean-variance analysis, which would not help with those negative returns securities that can be half (or more) of the sample in some cases? Or is it because it would cause overreactions in investors who would frantically sell/not buy securities and mutual funds with negative returns? There are other explanations. One of them is that authors may simply have been cautious to not overstate matters: "A number of stock returns are negative. If we ignore the effects of correlations, a negative expected return should imply a short position in the stock," Benninga (2008, p. 354). There is something else.

> **Observation 6**: Portfolio optimization presents serious numerical issues that have yet not been addressed in the literature.

And numerical issues may trouble the basis reduction method, as happened when optimizing for the Dow-Jones securities using monthly returns observations for the 6/1/2010 to 5-31/2010 period (Exhibit V, next section). The basis-reduction method wrongly eliminated one of the securities perhaps due to numerical troubles. The determinant of the variance-covariance matrix has a value of 1.37861E-82. This means the first significant digit, the one, appears at the decimal position 82, following 81 decimal zeros! This is very extreme numerical ill-conditioning warning of a very nearly singular matrix hardly worth inverting.

Three more comments close this subsection. 1) The basis reduction technique can be applied to the classical-Lagrangian optimization even is the starting return is not the tangent. Furthermore, fixing a given return in each step of the Lagrangian, instead of aiming at the tangent return, would provide the NSS solution for that required return. 2) For analytical purposes, note that signs in w follow signs in x in the algebraic formulation. 3) When the return on the portfolio is negative, the message seems to be that investors should not invest in the stock market for a while. In this case, torturing the data further (for instance, by applying the BR method) not only may not help but also is likely to provide misleading results.

III.3 Elton-Gruber-Padberg Algorithms

During the 1970's, two well-known Finance researchers, Edwin Elton and Martin Gruber teamed up with another well-known researcher in mathematical methods and optimization, Manfred Padberg, and developed index-based methods to optimize portfolios. Their collaborative efforts have and are still helping generations of students to learn a great deal about portfolio optimization (both single index, and mean-variance), and have motivated further important research, see Elton, Gruber, Brown, and Goetzmann (2010). Three items forming their research are particularly suited to our narrative. First, these authors show that maximizing the ratio $\theta = (rp - rf)/ stdp$ leads to model (7), Elton, Gruber, Brown, and Goetzmann (2010, 101 and ss.). Second, they also prove that ranking matters in portfolio optimization, which we will express as a critical observation because it supports one of the alternative optimization procedures we will cover shortly.

Observation 7: Ranking matters.

Finally, the Elton-Gruber-Padberg algorithms also illustrate that keeping all the securities and trying to draw the whole efficient frontier brings extra-items to the analysis that make it difficult to see what portfolio optimization does. One of the extra items is a constant correlation assumption for the mean-variance case.

III.4 Regression Approach

Being able to solve an optimal portfolio using regressions may have seemed like a dream to early researchers in portfolio optimization, but that is a reality for us. J. D. Jobson and R. M. Korkie's research has focused on the statistical foundations and characteristics of portfolio theory. On their way to test the efficiency of portfolio estimates, Jobson and Korkie (1983) unassumingly used a technique by which one can optimize a portfolio using regression --see their equation (19) on page 192 in the op. cit. They referred to this technique as "Least Square Estimation of the Efficient Set." Later on, Britten-Jones (199) fully developed this technique and applied it to estimate the sampling error of efficient portfolios (rp, varp). Tarrazo (2009b) provides a comparative analysis of regression and portfolio optimization and also shows that both mathematical specifications share the same set of homogeneous coordinates. This is very interesting for the purposes of this study because those coordinates embody arbitrage via rp = w'r for any valid specification of the optimization.

Exhibit III presents the application of the ordinary least squares regression method to portfolio optimization (POLS, for short). The model boils down to solving an unrestricted simultaneous equation system like (7). Using matrices, the beta coefficients are calculated with the "normal equations", $b = (X'X)^{-1} X'y$, and optimal weights with w = bi/sum(b), where X is the (n-by-k) returns matrix, y is a (n-by-1) vector of ones, and "n" and "k" are the number of observations and securities, respectively.

The POLS approach to portfolio analysis is a very important avenue of research --it can single-handedly disqualify portfolio optimization from practical endeavors. It also opens the "Econometrics" door to portfolio theory, as well as to other valuable vistas, see Scherer (2002). The POLS approach adds two critical observations to our treasure chest:

> **Observation 8**: Optimal Portfolios are also optimal predictors.
>
> **Observation 9**: Optimal tangent portfolios implement arbitrage and are optimal predictors

The eighth observation is important in both theoretical and practical terms: what else would an investor want to use other than an optimal predictor?

The ninth observation involves the two restrictions attached to the Lagrangian, which are subsumed in the tangent portfolio of the algebraic model. Tarrazo (2009b) shows that the estimated beta coefficients are defined in a Cartesian, unrestricted plane, but rebalancing makes them legitimate portfolio weights by restricting them to the sum(x) = 1 plane. Arbitrage requires that w' r = rp, and this happens in that plane because the optimal weight of the portfolio equals one, wp = sum(w) = 1. Homogeneous coordinates link the unrestricted Cartesian plane to the restricted sum(x) = 1 (barycentric) plane, where arbitrage is effective.

No-short sales solutions using POLS are easy to obtain, either by basis reduction of the corresponding returns columns, or simply by using software that provides nonnegative ordinary least squares (OLS) routines –for example, the "lsqnonneg" function in MATLAB. The reader is referred to the references provided in this section for further details on the POLS method.

Exhibit III. Least Squares Portfolio Optimization and Basis Reduction.

X'X							X'y
0.157393	0.064978	0.101543	0.08692	0.082453	0.119554		1.014642
0.064978	0.221245	0.146545	0.076243	0.014701	0.134815		0.588078
0.101543	0.146545	0.299371	0.092845	0.097178	0.220324		0.651069
0.08692	0.076243	0.092845	0.146425	0.064696	0.16759		0.328404
0.082453	0.014701	0.097178	0.064696	0.254058	0.185373		-0.42537
0.119554	0.134815	0.220324	0.16759	0.185373	0.620534		-0.68863

	SS, tangent		BR1		BR2, NSS tangent	
	bi	wi	bi	wi	bi	wi
MCD	8.679794	164.47%	6.312622	91.00%	6.087297	87.49%
IBM	0.643612	12.20%	1.156895	16.68%	0.870238	12.51%
HPQ	1.920975	36.40%	-0.53269	-7.68%		
KO	-0.002109	-0.04%				
PFE	-3.367643	-63.81%				
GE	-2.597308	-49.22%				
	5.277321	1	6.936828	1	6.957535	1
mdeterm	3.05E-05		0.005433		0.0306	

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.477082
R Square	0.227607
Adjusted R Square	0.137571
Standard Error	0.926398
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	13.65644	2.276074	2.652105	0.025259
Residual	54	46.34356	0.858214		
Total	60	60			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
MCD	8.679794	3.080035	2.818083	0.006737	2.504697	14.85489	2.504697	14.85489
IBM	0.643612	2.554333	0.251969	0.802022	-4.47752	5.764741	-4.47752	5.764741
HPQ	1.920975	2.363257	0.812851	0.419873	-2.81707	6.659019	-2.81707	6.659019
KO	-0.00211	3.366018	-0.00063	0.999502	-6.75057	6.74635	-6.75057	6.74635
PFE	-3.36764	2.254762	-1.49357	0.141109	-7.88817	1.152882	-7.88817	1.152882
GE	-2.59731	1.599059	-1.62427	0.110141	-5.80323	0.608612	-5.80323	0.608612

Exhibit IV. Selective Basis Addition.

	MCD	IBM	HPQ	KO	PFE	GE		Returns
MCD	0.002337	0.000917	0.001509	0.001356	0.001494	0.002187		0.016911
IBM	0.000917	0.003591	0.002336	0.001217	0.000315	0.002359		0.009801
HPQ	0.001509	0.002336	0.004872	0.001488	0.001697	0.003797		0.010851
KO	0.001356	0.001217	0.001488	0.00241	0.001117	0.002856		0.005473
PFE	0.001494	0.000315	0.001697	0.001117	0.004184	0.003008		-0.00709
GE	0.002187	0.002359	0.003797	0.002856	0.003008	0.010211		-0.01148

	MCD	IBM			Returns	
MCD	0.002337	0.000917			0.016911	MCD
IBM	0.000917	0.003591			0.009801	IBM
	xi	wi				
MCD	6.850974	0.874922				
IBM	0.979413	0.125078				
	7.830387	1				
mdeterm	7.55E-06					

	MCD	IBM	HPQ		Returns	
MCD	0.002337	0.000917	0.001509		0.016911	MCD
IBM	0.000917	0.003591	0.002336		0.009801	IBM
HPQ	0.001509	0.002336	0.004872		0.010851	HPQ
	xi	wi				
MCD	7.111288	0.910016				
IBM	1.303265	0.166776				
HPQ	-0.60008	-0.07679				
	7.814468	1				
mdeterm	2.23E-08					

	MCD	IBM	KO		Returns	
MCD	0.002337	0.000917	0.001356		0.016911	MCD
IBM	0.000917	0.003591	0.001217		0.009801	IBM
KO	0.001356	0.001217	0.00241		0.005473	KO
	xi	wi				
MCD	8.535015	1.247009				
IBM	1.697506	0.248014		Termination		
KO	-3.38813	-0.49502		No more ri's > 0		
	6.844391	1				
mdeterm	1.12E-08					

III.5 Basis Addition

Exhibit IV shows the last alternative optimization method in this study. It was introduced by Tarrazo (2009a), who called it the "Selective Basis Addition" (SBA) method. It is the logical counterpart to the basis reduction (BR) method which, starting from the whole set of securities "peels off" the unwanted ones according to the sign of their weights. In contrast, the SBA method takes advantage of the Elton-Gruber-Padberg ranking strategy and starts at the top of the sample, which is the optimal portfolio formed by the top two securities when ranked according the individual return-to-risk ($r_i/pstd_i$) criteria. Samuelson (1967) showed that the asset with the largest mean will always be included in an optimal portfolio, and Markowitz (1959) advised to avoid extreme outcomes, like that of investing only in one asset. Two, therefore, is the minimal size portfolio to start the SBA algorithm.

Securities with negative returns are, of course, summarily discarded before starting the SBA, and each security with positive returns is "tried out", one at a time. The process is terminated when two negative optimal weights are found, one after the other, and after trying out all the securities with positive weights. Those two securities are to be discarded. As explained in Tarrazo (op. cit.), this stopping rule is borrowed from graph theory. A balanced graph is one where all the signs, which express relationships among the individual elements, have the same sign. In the portfolio case, the chain is broken when one negative sign comes up, and two negative optimal weights should signal that there are no more optimal securities left.

Tarrazo (op. cit.) advises to check each security with positive returns because at some point the risk-to-return ranking may not provide strong signals. This, however, may not be necessary. The fact that we can add and subtract basis on our way to the optimal solution is a characteristic of "separable" programming, which uses "ordered sets." Another characteristic of separable programming is the "restrictive basis entry." Because the ordering of the sets matters, there are expectations concerning what the entrants could and would do to the optimization. The following paragraph from the Mathematical Programming Glossary of the INFORMS Computing Society is self-explanatory:

> "A common rule arises in separable programming, which uses specially ordered sets: a group of non-negative variables must sum to 1 such that at most two variables are positive, and if two are positive, they must be adjacent. For example, suppose the variables are (x1,x2,x3). Then, it is feasible to have (.5,.5,0) and (0,.2,.8), but it is not feasible to have (.5,0,.5) or (.2,.2,.6). In this case the rule is not to permit a variable to enter the basis unless it can do so without violating the adjacency requirement. For example, if x1 is currently basic, x3 would not be considered for entry… Another restricted entry rule pertains to the delta form of separable programming (plus other applications): Do not admit a variable into the basis unless its predecessor variables are at their upper bound."
> (INFORMS Computing Society, Mathematical Programming Glossary)

"Ordered sets" are used in discrete optimization, see Beale and Tomlin (1970) and Beale and Forrest (1976). The SBA is a case of discrete optimization because it adds one security at a time. The examples provided by Tarrazo (2009a) conform the expectations set for the variables in the above quotation-paragraph. Beale, as in E. M. L. Beale, is the author of seminal contributions to the literature on quadratic programming we reviewed in the first section of this study.

The SBA method contributes the last three observations:

> **Observation 10**: Rank not only matters, it governs the optimization. The individual security return-to-risk (r_i/$pstd_i$) suffices to exploit rank-based ordering in the optimization.
> **Observation 11**: The role played by variance and covariance.

Preliminary experience with the SBA seems to indicate that individual variances are more important than is commonly thought. This is confirmed by Tarrazo's (2008) analysis of positive weights in the efficient frontier. The image we have of the role of covariances in optimization is severely distorted because it is often studied in a context where the investor is supposed to randomly select and add equally weighted securities ad infinitum, which is something an actual investor (individual or institutional) would never, ever, do. The threat of liability incurred by the potential losses is enough for institutional investors to forget about that idea.

> **Observation 12**: The so-called minimum-variance portfolio is the *summum malum* –the worst possible portfolio-- to choose among the presumably "efficient ones."

Even precluding negative weights (NSS), the minimum-variance portfolio it may include securities with negative returns, and many of those securities with positive returns that do not hold their own when maximizing the return-to risk ratio of the portfolio. The weakness in their returns is arbitrarily hidden from the optimizer's reach. This amounts to risky data manipulation and is likely to result in qualitatively dismal portfolios. In other words, what we call the minimum variance portfolio is unlikely to have the minimum risk.

Our review and analysis of optimization methods has come to an end. Further information on optimization can be found in Nocedal and Wright (1999, Chapter 16), who provide an excellent update on modern approaches to quadratic programming in portfolio optimization. Keep in mind, however, that in many cases the method itself veils what the optimization does, e.g., "interior methods." In contrast, our strategy has been to learn as much as possible about the problem to simplify the optimization to the maximum. In other words, we learned from QP and other methods to come up with more transparent approaches. Further, we proceeded by taking advantage of valuable leads that in several cases illustrate modern approaches to optimization but were not taken advantage of when first proposed. For example, the basis reduction method makes the best of the restricted basis entry in Wolfe's (1959) algorithm. Both the basis reduction and the selective basis addition put to good use Beale's technique of allowing the tableau (and, therefore, the basis of the problem) to change. Our emphasis on the unrestricted form has some connections to Lemke's method. It also implements Beale's suggestion of employing the Newton-Raphson method. In the quadratic case, the Newton-Raphson method is so efficient that the optimal is reached in one step ($x^* = A^{-1} b$). But our persistence in using simultaneous equation systems, the algebraic method, pre-processing the data by ranking, and trying out diverse heuristics belongs squarely to modern optimization, as noted at the beginning of this section.

Exhibit V. Optimizations, Dow-Jones Industrials Average.

	6/1/2010-5/31/2005			4/1/2008-3/31/2003			11/3/2003-12/01/1998		
		sorted	NSS		sorted	NSS		sorted	NSS
	Ticker	ri/stdi	tangent	Ticker	ri/stdi	tangent	Ticker	ri/stdi	tangent
1	**MCD**	0.349792	**0.874922**	**MCD**	0.422509	**0.203917**	**MMM**	0.261209	**0.820107**
2	**IBM**	0.163552	**0.125078**	**UTX**	0.379721	**0.084**	**C**	0.154224	**0.160004**
3	HPQ	0.155465		**CVX**	0.37362	**0.116184**	CAT	0.106454	0
4	KO	0.111483		**XOM**	0.32892	**0.011961**	UTX	0.096706	0
5	TRV	0.105293		**CAT**	0.322833	**0.086492**	AA	0.095815	0
6	CVX	0.09833		**BA**	0.317197	**0.112493**	AXP	0.074469	0
7	UTX	0.095426		HPQ	0.287006	0	**WMT**	0.068275	**0.002498**
8	PG	0.084841		**T**	0.256979	**0.048972**	JNJ	0.063278	0
9	T	0.077528		**PG**	0.241008	**0.02235**	XOM	0.055344	0
10	MRK	0.072305		KO	0.210102	0	BA	0.038652	0
11	DIS	0.067789		**JPM**	0.204281	**0.055069**	MO	0.036374	0
12	MMM	0.058056		DIS	0.190007	0	**PG**	0.031413	**0.017391**
13	CAT	0.051682		GE	0.17703	0	HPQ	0.020182	0
14	JPM	0.030403		AXP	0.146239	0	INTC	0.014146	0
15	WMT	0.029817		**JNJ**	0.135872	**0.150697**	IBM	0.002148	0
16	XOM	0.029685		**BAC**	0.131659	**0.01768**	IP	-0.00982	
17	VZ	0.026016		AA	0.129226	0	GE	-0.01169	
18	CSCO	0.022831		IBM	0.122984	0	HD	-0.01277	
19	JNJ	0.016389		MMM	0.116577	0	DD	-0.02044	
20	KFT	0.014402		DD	0.095524	0	GM	-0.02099	
21	BA	0.011325		VZ	0.083618	0	MSFT	-0.0364	
22	MSFT	0.001698		MSFT	0.082953	0	HON	-0.03644	
23	DD	-0.00291		INTC	0.050069	0	JPM	-0.03657	
24	AXP	-0.00955		HD	0.032792	0	DIS	-0.03945	
25	INTC	-0.03481		**WMT**	0.009282	**0.090185**	KO	-0.06381	
26	HD	-0.04247		MRK	-0.03431	0	MCD	-0.06783	
27	BAC	-0.08715		AIG	-0.04532	0	MRK	-0.08187	
28	AA	-0.09757		GM	-0.05196	0	SBC	-0.10922	
29	PFE	-0.1096		PFE	-0.06166	0	EK	-0.12969	
30	GE	-0.11358		C	-0.08424	0	T	-0.26613	
monthly	rp		0.016021			0.014226			0.016352
	pvar		0.002046			0.000575			0.003744
	pstd		0.045233			0.023975			0.061189
	rp/pstd		0.354195			0.593362			0.267228
	rp/pvar		7.830387			24.74924			4.367219
yearly									
	rp		19.23%			17.07%			19.62%
	pvar		0.024553			0.006898			0.04493
	pstd		15.67%			8.31%			21.20%
	rp/pstd		1.226968			2.055466			0.925704
	rp/pvar		7.830387			24.74924			4.367219
mdeterm		1.38E-82	7.55E-06		-2.44E-85	1.80E-34		9.31E-71	1.06E-09
emax = norm (k)		0.1072	0.0041		0.0242	0.0087		0.108146	0.01127
emin		0.000165	0.001864		5.16E-05	0.00035		0.000401	0.002958
cond(k) = emax/emin		651.6525	2.1991		469.2584	24.8673		269.7937	3.809661

III Additional investing implications

It is appropriate to leave the confines of our example and observe more realistically sized portfolio optimizations. This is the objective of the numbers presented in Exhibit V. Several issues are worth noting if we are to learn from the material presented in the previous sections, such as the following: 1) the number of stocks in the NSS portfolio, 2) the number of securities with negative returns, and 3) those securities with positive returns but excluded from the tangent portfolio.

The optimizations confirm what has been observed numerous times. The “best” and presumably most reliable portfolio is disappointingly composed of an unrealistically small number of securities. This happens in the presence of both relatively numerous securities with negative returns and those with relatively few. Even the optimization that appears better balanced is not so upon further inspection. There are twelve optimal securities, but half of them (MCD, JNJ, CVX, BA, WMT, and CAT) take on 75.99% of the total portfolio. Note that our starting sample is the Dow-Jones Industrial Average which is not only well within the recommended optimal number of initial securities by some authors, but exhibits a correlation of 98.18% with S&P 500 for the 6/1/2010-5/31/2005 period.

With respect to the relative efficiency of each alternative method, it is noteworthy that the presence of many securities with negative returns seems to favor both the basis reduction (BR) and the selective basis addition (SBA); the acute concentration of the optimal portfolio in the top securities according to the r_i/std_i ranking definitely favors the SBA model, which also benefits from rejecting the securities with negative returns. More importantly, the optimizations support the critical observations we derive when studying each optimization alternative and have incidence in several strands of current research:

a) Restrictions. The analysis presented supports studying the results from optimizing with no restrictions other than that of no-short sales first. Restrictions that bring the resulting portfolio closer to the tangent would prove to be beneficial in the sense that they reduce the securities that are too weak to be part of the highest rp/pvar portfolio, even if they have positive returns. The distortion caused by the usual required return restriction may be particularly hard to assess.
b) Our analysis also discourages other manipulation of the data such as resampling which, rather than lessening the limitations of mean-variance, is likely to deform investing in at least two manners: First, by imposing man-made uncertainty (Monte Carlo trials) into actual investing positions); Second, by ignoring what the data may say about the actual economic reality and security characteristics.
c) As noted earlier, the minimum variance portfolio may nominally have the lowest variance but it is not likely to be the one with the minimum risk one. This happens not because it excludes returns considerations, as is sometimes mistakenly noted, but because even the NSS minimum variance optimization takes any returns (including negative ones) without discriminating among them.
d) Our critical observations point towards paying more attention to the economic logic of the problem, for example, how the economy may have affected the sample, and how the

characteristics of firms (fundamentals) may be related to the ranking heuristic r_i/std_i. For individual investors, who do not have money to invest in many securities, such heuristic can anticipate how the security would fare in an optimization. This investor could also consider simply holding the top “k” securities according to such ranking.

What are the challenges that investors face when trying to mean-variance optimization? Mainly these three:

1. The time window selected to collect the sample data is critical for the performance of the portfolio and progress is difficult in this area because of conflicts between strengthening some of the statistical properties of the estimates and capturing a sample representative of the future, expected, conditions. For example, Jobson and Korkie (1981) note that, “(T)he number of historical observations of monthly returns required to give reasonably unbiased estimates of the optimal risk and return estimates is at least two hundred … The traditional Markowitz procedure for predicting the optimal risk-return is extremely poor with conventional sample sizes of four to seven years. Increasing sample sizes to the requirement suggested here is perhaps untenable because of changing market conditions over such extended periods,” (op. cit., p. 72).
2. The presence of extreme weights. This happens not because the optimizer is malfunctioning but because mean-variance optimization is ruthless in letting the winner take all. At the time of this writing, place companies like Google, or Apple in a small portfolio and you will see them taking most of the optimal weight and “ruining” the optimization, in the sense that the rest of optimal weights are deformed by the superstars.
3. In its present form, portfolio analysis suffers from serious numerical problems. And adding more observations and more securities only likely to make the numerical problem worse. The central matrix in the analysis (covariance) quickly becomes unreliable for inversion: a very small change caused by inaccuracy of the estimates for example is likely to have enormous impact on the optimal weights. Econometric results are “treated” for condition numbers around 10, and around 20 and above are definitely considered to be in the “danger level.” The ones in Exhibit V are 651.6525, 469.25, and 269.7937 for the “full matrices” (the ones with 30 securities). The ones for the tangent portfolios are much better numerically: 2.1991, 24.8673, and 3.809661, respectively. Just as the least squares approach opens the “econometric” landscape to portfolio optimization, the algebraic model also opens the numerical analysis area to portfolio optimization. Tarrazo (2010) shows that the algebraic approach provides key information that is not obtainable in the regular optimization. Part of this information concerns the structural strength of the portfolio. For example, the larger the portfolio, the lower its structural strength, as we have seen in our applications. This is bad news for the minimum variance portfolios which, in addition to being qualitatively deficient, may also have poor numerical properties. In general, structural analysis is a missing piece in portfolio construction: one seems to be careful of not putting all the eggs in one basket but completely fails to consider the strength of the basket, and whether it is worthwhile to put anything in it.

CONCLUDING COMMENTS

The question opening this study returns: Why do Finance professors (and other investors) not use portfolio optimization? In our opinion, lack of knowledge about what portfolio theory does exactly may be part of the problem.

Our response has been to open the portfolio optimization engine box and to look carefully inside. We have made a number of critical observations that have clear investing implications and are likely to help assess the practical value of portfolio analysis. The cumulative effect of the key observations derived from the analysis of portfolio optimization is promising. Further progress can be achieved through a two-pronged approach. First, the optimization must be kept as simple as possible, avoiding any data manipulation. This strategy should help isolate the three major challenges faced by investors who want to use mean-variance analysis: the performance reliability of the sample selected; the "winner takes all" characteristic of the optimization; and the rather severe numerical weaknesses.

Second, many of the roads explored lead to small portfolios (e.g., the Dow-Jones Industrials Average Index and groupings of similar size). There is no telling what type of jumbled signals can originate from portfolios of hundreds of securities being optimized under arbitrary restrictions. Small portfolios have the best numerical and statistical properties. In addition, small portfolios are the only ones that can be followed with enough precision to ascertain the effects that economic, market, and individual conditions have on the optimal weights. This means that information can be used to complement the results from optimizations and used to manage risk. The horizon thus opened is less theoretical and more practical, and the effort is likely to lead beyond statistical indicators and monthly returns into production and other primary indicators of a firm's performance.

Small portfolios may serve individual investors well. That was the inspiration that created mean-variance analysis to begin with, and there is certain urgency in helping small (household) investors today. The overwhelming presence of mutual funds as the only game these investors can play, the concentration of trading among powerful institutional investors, and the sharp episodes of volatility may be turning the original goal of portfolio theory –that is, helping household investors to purchase individual stocks—into a romantic chimera of times gone by.

REFERENCES

Ball, M. (2011). Heuristics Based on Mathematical Programming. *Surveys in Operations Research and Management Science*, Vol. 16, No. 1, 21-38.

Beale, E. (1955). On Minimizing A Convex Function Subject to Linear Inequalities. *Journal of the Royal Statistical Society*. Series B (Methodological), Vol. 17, No. 2, pp. 173-184.

Beale, E. M. L. (1959). On Quadratic Programming. *Naval Research Logistics Quarterly*, 6: 227–243.

Beale, E. (1967). Numerical Methods, Chapter VII, pp. 134-205, in J. Abadie eds. *Nonlinear Programming*. John Wiley & Sons, Inc., New York, North-Holland Publishing Company, Amsterdam.

Beale, E., and Forrest, J. (1976). Global Optimization Using Special Ordered Sets. *Mathematical Programming*, 10(1):52–69.

Beale, E., and Tomlin, J. (1970). Special Facilities in a General Mathematical Programming System for Nonconvex Problems Using Ordered Sets of Variables. In J. Lawrence, editor, *Operational Research 69*, pages 447–454. Tavistock Publishing, London.

Benjelloun, H. (2010). Evans and Archer –Forty Years Later. *Investment Management and Financial Innovations*, Volume 7, Issue 1, 98-104.

Benninga, S.(2008). *Financial Modeling*. 3rd ed., The MIT Press, Cambridge, Massachusetts.

Boot, J. (1964). *Quadratic Programming*. North Holland Publishing Co., Amsterdam.

Boot, J. (1963). Binding Constraint Procedures of Quadratic Programming.*Econometrica*, Vol. 31, No. 3, July 1963, 464-498.

Boot, J. (1962). On Trivial and Binding Constraints in Programming Problems. *Management Science*, Vol. 8, 419-441.

Boscaljon, B., Filbeck, G., and Ho, C. (2005). How Many Stocks Are Required for a Well-Diversified Portfolio? *Advances in Financial Education*, Fall, Vol. 3, 60-71.

Britten-Jones, M. (1999). The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights. *The Journal of Finance*, April 1999, Vol. LIV, No. 2, 655-671.

Chow, G., Jacquier, E., Kritzman, M., and Lowry, K. (1999). Optimal Portfolios in Good Times and Bad. *Financial Analyst Journal*, May/June, Vol. 55, No. 3, 65-73.

Coger, K., and Ruland, P. (1983). A Goal Seeking Investment Model. *Management Science*, Vol. 29, No. 9, September, 1027-1036.

Computing Society INFORMS. (2010). *Mathematical Programming Glossary*. Last accessed March 15, 2010. http://glossary.computing.society.informs.org/second.php?page=S.html#SOS

Constantinides, G. and A. Malliaris. (1995). Portfolio Theory, in Jarrow, R., Maksimovic, V., and W. Ziemba, eds.: *Handbook in OR & MS*, Vol. 9. Chapter 1, pp. 1-30.Elsevier Science, B.V., Amsterdam, The Netherlands.

Dickinson, J. (1974). The Reliability of Estimation Procedures in Portfolio Analysis. *Journal of Financial and Quantitative Analysis*, June, Vol. 9, 447-462.

Domian, D., Louton, D., and Racine, M. (2007). Diversification in Portfolios of Individual Stocks: 100 Stocks Are Not Enough. *The Financial Review*, November, Vol. 42, No. 4, 557-570.

Doran, J., and Wright, C. (2010). What Really Matters When Buying and Selling Stocks: A Survey of Finance Professors' Investing Behavior. *Advances in Financial Education*, Vol. 8, Issues 1 & 2, 35-61.

Ellerman, D. (1984). Arbitrage Theory: A Mathematical Introduction. *SIAM Review* , Vol. 26, No. 2, April, 241-26.

Ellerman, D. (1990). An Arbitrage Interpretation of Classical Optimization. *Metroeconomica*, Volume 41, Issue 3, October, 259–276,

Elton, E. and Gruber, M., Brown, S., and Goetzman, W. (2010). *Modern Portfolio Theory and Investment Analysis*. 8th. ed., Wiley, New York.

Evans, J., and Archer, S. (1968). Diversification and the Reduction of Dispersion: An Empirical Analysis. *Journal of Finance*, 23, 761-767.

Fabozzi, F., Gupta, F., and Markowitz, H. (2002). The Legacy of Modern Portfolio Theory. *The Journal of Investing*, Fall, Vol. 1, No. 3, 7-22.

Francis, J. and Archer, S. (1971). *Portfolio Analysis*. Prentice Hall, Englewood Cliffs, New Jersey.

Geoffrion, A. (1976). The Purpose of Mathematical Programming Is Insight, Not Numbers. *Interfaces*, Vol. 7, No. 1, Part 1 of Two, November, 81-92.

Hsu, C., and Wei, J. (2003). Stock Diversification in the U.S. Equity Market. *B>Quest*, 2003: http://www.westga.edu/~bquest/2003/diversify.htm

Jaganathan, R., and Kocherlakota, M. (1996). Why Should Older People Invest Less in Stocks than Younger People? *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 20, No. 3, Summer, 11-23.

Jarrow, R. (1988). *Finance Theory*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Jeffrey, R. (1984). A New Paradigm for Portfolio Risk. *The Journal of Portfolio Management*, Fall, Vol. 1, No. 1, 33-40.

Jobson, J., and Korkie, B. (1981). Putting Markowitz Theory to Work. *The Journal of Portfolio Management*, Summer, Vol. 7, No. 4, 70-74.

Jobson, J., and Korkie, R. (1983). Statistical Inference in a Two-Parameter Portfolio Theory with Multiple Regression Software. *Journal of Financial and Quantitative Analysis*, Vol. 18, No. 2, June 1983, 189-197.

Kwan, C., and Yuan, Y. (1993). Optimal portfolio selection without short sales under the full-information covariance structure: A pedagogic consideration. *Journal of Economics and Business*, Volume 45, Issue 1, February, 91–98.

Lee, H., and Hanna, S. (1995). Investment Portfolios and Human Wealth. *Financial Counseling and Planning*, Vol. 6, 147-152.

Lee, W. (1990). Diversification and Time: Do Investment Horizons Matter? *The Journal of Portfolio Management*, vol. 16, no. 3 (Spring), 21-26.

Lemke, C. (1962). A Method of Solutions for Quadratic Programs. *Management Science*, Vol. 8, No. 4, July, 442-453.

Luenberger, D. (1984). *Linear and Nonlinear Programming*." 2nd ed., Addison-Wesley Publishing Company, Reading, Massachusetts.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, Vol. VII, No. 1, 77-91.

Markowitz, H. (1956). The Optimization of a Quadratic Function Subject to Linear Constraints. *Naval Research Logistics Quarterly*, 3: 111–133.

Markowitz, H. (1959). *Portfolio Selection, Efficient Diversification of Investments*. New York, John Wiley and Sons.

Markowitz, M. (1999). The Early History of Portfolio Theory. *Financial Analyst Journal*, Vol. 55, No. 4, July-August, 5-16.

Martellini, L, Urosevic, B. (2006). Static Mean-Variance Analysis with Uncertain Time Horizon. *Management Science*, Vol. 52, No. 6, June, 955-964.

Martin, A. (1955). Mathematical Programming of Portfolio Selections. *Management Science*, January, Vol. 1, No. 2, 152-166.

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*, 34, 768-783.

Nemhauser, G., and Wolsey, L. (1988). *Integer and Combinatorial Optimization*. John Wiley and Sons, New York.

Newbould, G. and Poon, P. (1993). The Minimum Number of Stocks Needed for Diversification. *Financial Practice and Education*, Fall, Vol. 3, No. 2, 85-87.

Nocedal, J., and S. Wright. (1999). *Numerical Optimization*. Springer Verlag, New York.

Pardalos, P, and Rosen, J. (1987). *Constrained Global Optimization: Algorithms and Applications*. Lecture Notes in Computer Science 268, Springer Verlag, New York.

Renshaw, E. (1977). Short Selling and Financial Arbitrage. *Financial Analyst Journal*, January-February, Vol. 33, No. 1, 58-65.

Roy, A. (1952). Safety First and the Holding of Assets. *Econometrica*, July, 20, 431-49.

Rubinstein, M. (2002). Markowitz's 'Portfolio Selection': A Fifty-Year Retrospective. *The Journal of Finance*, June, Vol. LVII, No. 3, 1041-1045.

Rubinstein, M. (2006). *A History of the Theory of Investments: My Annotated Bibliography*. Wiley, Hoboken, New Jersey.

Samuelson, P. (1967). A General Proof that Diversification Pays. *Journal of Financial and Quantitative Analysis*, Vol. 2, No. 1, 1-13.

Samuelson, P. (1983). *Foundations of Economic Analysis*. Harvard University Press, enlarged edition, Boston, Massachusetts. (Originally published in 1947)

Scherer, B. (2002). Portfolio Resampling: Review and Critique. *Financial Analyst Journal*, November-December, 98-109.

Sharpe, W. (1991). Capital Asset Prices with and without Negative Holdings. *The Journal of Finance*, June, Vol. XLVI, No. 2, pp. 489-509.

Shawky, H., and Smith, D. (2005). Optimal Number of Stocks Holdings in Mutual Fund Portfolios based on Market Performance. The Financial Review, 40, 481-495.

Silver, E. (2004). An Overview of Heuristic Solution Methods. *Journal of the Operational Research Society*, Volume 55, Number 9, September, pp. 936-956.

Simmons, D. (1975). *Nonlinear Programming for Operations Research*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Statman, M. (1987). How Many Stocks Make a Diversified Portfolio? *Journal of Financial and Quantitative Analysis*, Vol. 22, No. 3, 353-364.

Steinbach, M. (2001). Markowitz Revisited: Mean-Variance Models in Financial Portfolio Analysis. *SIAM Review*, Vol. 43, No. 1, 31-85.

Tarrazo, M. (2000). Mathematical Programming and Portfolio Optimizations: A Clarification. *Advances in Financial Planning and Forecasting*, Vol. 9, 31–55.

Tarrazo, M. (2009a). Identifying Securities to Buy: The Heuristic ri/stdi. Research in Finance, Vol. 25, 229-268.

Tarrazo, M. (2009b). On Portfolios and Regressions. *The Journal of Financial Education*, Vol. 35, Fall, 56-74.

Tarrazo, M. (2008). Positive Optimal Weights on the Efficient Frontier: Conditions, Causes, and Implications. *The International Journal of Finance,* Vol. 20, No. 2, 4743-4774.

Tarrazo, M. (2010). Algebraic Analysis of Optimal Portfolios. *Advances in Investment Analysis and Portfolio Management*, Vol. 4, 113-152.

Tarrazo, M., and Murray, W. (2004). Teaching What We Know about Asset Allocation. *Advances in Financial Education*, Spring, Vol. 2, 77-103.

Theil, H., and Van de Panne, C. (1960). Quadratic Programming as an Extension of Classical Quadratic Maximization. *Management Science*, Vol. 7, No. 1, October, 1–20.

Tobin, J. (1958). Liquidity Preference as Behavior Towards Risk. *Review of Economic Studies*, 25, 65-86.

Wolfe, P. (1959). The Simplex Method for Quadratic Programming. *Econometrica*, Vol. 27, No. 3, July 1959, 382-398.