The Early History of Portfolio Theory: 1600-1960

Author(s): Harry M. Markowitz

The Early History of Portfolio Theory: 1600–1960

Harry M. Markowitz

*D*iversification of investments was a well-established practice long before I published my paper on portfolio selection in 1952. For example, A. Wiesenberger's annual reports in *Investment Companies* prior to 1952 (beginning 1941) showed that these firms held large numbers of securities. They were neither the first to provide diversification for their customers (they were modeled on the investment trusts of Scotland and England, which began in the middle of the 19th century), nor was diversification new then. In the *Merchant of Venice*, Shakespeare has the merchant Antonio say:

> My ventures are not in one bottom trusted,
> Nor to one place; nor is my whole estate
> Upon the fortune of this present year;
> Therefore, my merchandise makes me not sad.
>> Act I, Scene 1

Clearly, Shakespeare not only knew about diversification but, at an intuitive level, understood covariance.

What was lacking prior to 1952 was an adequate *theory* of investment that covered the effects of diversification when risks are correlated, distinguished between efficient and inefficient portfolios, and analyzed risk–return trade-offs on the portfolio as a whole. This article traces the development of portfolio theory in the 1950s (including the contributions of A.D. Roy, James Tobin, and me) and compares it with theory prior to 1950 (including the contributions of J.R. Hicks, J. Marschak, J.B. Williams, and D.H. Leavens).

Portfolio Theory: 1952

On the basis of Markowitz (1952), I am often called the father of modern portfolio theory (MPT), but Roy (1952) can claim an equal share of this honor. This section summarizes the contributions of both.

My 1952 article on portfolio selection proposed expected (mean) return, E, and variance of return, V, of the portfolio as a whole as criteria for portfolio selection, both as a possible hypothesis about actual behavior and as a maxim for how investors ought to act. The article assumed that "beliefs" or projections about securities follow the same probability rules that random variables obey. From this assumption, it follows that (1) the expected return on the portfolio is a weighted average of the expected returns on individual securities and (2) the variance of return on the portfolio is a particular function of the variances of, and the covariances between, securities and their weights in the portfolio.

Markowitz (1952) distinguished between efficient and inefficient portfolios. Subsequently, someone aptly coined the phrase "efficient frontier" for what I referred to as the "set of efficient mean–variance combinations." I had proposed that means, variances, and covariances of securities be estimated by a combination of statistical analysis and security analyst judgment. From these estimates, the set of efficient mean–variance combinations can be derived and presented to the investor for choice of the desired risk–return combination. I used geometrical analyses of three- and four-security examples to illustrate properties of efficient sets, assuming nonnegative investments subject to a budget constraint. In particular, I showed in the 1952 article that the set of efficient portfolios is piecewise linear (made up of connected straight lines) and the set of efficient mean–variance combinations is piecewise parabolic.

Roy also proposed making choices on the basis of mean and variance of the portfolio as a whole. Specifically, he proposed choosing the portfolio that maximizes portfolio $(E - d)/\sigma$, where d is a fixed (for the analysis) disastrous return and σ is standard deviation of return. Roy's formula for the variance of the portfolio, like the one I presented, included the covariances of returns among securities. The chief differences between the Roy analysis and my analysis were that (1) mine required nonnegative investments whereas Roy's allowed the amount invested in any security to be positive or negative and (2) I proposed allowing the investor to choose a desired portfolio from the efficient frontier whereas Roy recommended choice of a specific portfolio.

Harry M. Markowitz is president of Harry Markowitz Company.

Comparing the two articles, one might wonder why I got a Nobel Prize for mine and Roy did not for his. Perhaps the reason had to do with the differences described in the preceding paragraph, but the more likely reason was visibility to the Nobel Committee in 1990. Roy's 1952 article was his first and last article in finance. He made this one tremendous contribution and then disappeared from the field, whereas I wrote two books (Markowitz 1959; Markowitz 1987) and an assortment of articles in the field. Thus, by 1990, I was still active and Roy may have vanished from the Nobel Committee's radar screen.

Problems with Markowitz (1952). I am tempted to include a disclaimer when I send requested copies of Markowitz (1952) that warns the reader that the 1952 piece should be considered only a historical document—not a reflection of my current views about portfolio theory. There are at least four reasons for such a warning. The first two are two technical errors described in this section. A third is that, although the article noted that the same portfolios that minimize standard deviation for given E also minimize variance for given E, it failed to point out that standard deviation (rather than variance) is the intuitively meaningful measure of dispersion. For example, "Tchebychev's inequality" says that 75 percent of any probability distribution lies between the mean and ±2 *standard deviations*—not two variances. Finally, the most serious differences between Markowitz (1952) and the views I now hold concern questions about "why mean and variance?" and "mean and variance of what?". The views expressed in Markowitz (1952) were held by me very briefly. Those expressed in Markowitz (1959) have been held by me virtually unchanged since about 1955. I will discuss these views in the section on Markowitz (1959).

As for the technical errors: First, it has been known since Markowitz (1956) that variance, V, is a strictly convex function of expected return among efficient EV combinations. Markowitz (1952) explained, correctly, that the curve is piecewise parabolic. Figure 6 of Markowitz (1952) showed two such parabola segments meeting at a point. The problem with the figure is that its parabolas meet in such a way that the resulting curve is not convex. This cannot happen. Second, Figure 1 in Markowitz (1952) was supposed to portray the set of all feasible EV combinations. In particular, it showed the "inefficient border," with *maximum* V for a given E, as a concave function. This is also an error. Since my 1956 article, we know that the curve relating maximum V for a given E is neither concave nor convex (see Markowitz 1987, Chapter 10, for a description of possibilities).

The "General" Portfolio Selection Problem. For the case in which one and only one feasible portfolio minimizes variance among portfolios with any given feasible expected return, Markowitz (1952) illustrated that the set of efficient portfolios is piecewise linear. It may be traced out by starting with the unique point (portfolio) with minimum feasible variance, moving in a straight line from this point, then perhaps, after some distance, moving along a different straight line, and so on, until the efficient portfolio with maximum expected return is reached.[1] Note that we are not discussing here the shape of efficient mean–variance combinations or the shape of efficient mean–standard deviation combinations. Rather, we are discussing the shape of the set of efficient portfolios in "portfolio space."

The set of portfolios described in the preceding paragraph is *not* a piecewise linear *approximation* to the problem; rather, the exact solution is itself piecewise linear. The points (portfolios) at which the successive linear pieces meet are called "corner portfolios" because the efficient set turns a corner and heads in a new direction at each such point. The starting and ending points (with, respectively, minimum variance and maximum mean) are also called corner portfolios.

Markowitz (1952) did not present the formulas for the straight lines that make up the set of efficient portfolios. These formulas were supplied in Markowitz (1956), but Markowitz (1956) solved a much more general problem than discussed in Markowitz (1952). A portfolio in Markowitz (1952) was considered feasible ("legitimate") if it satisfied one equation (the budget constraint) and its values (investments) were not negative. Markowitz (1956), however, solved the (single-period mean–variance) portfolio selection problem for a wide variety of possible feasible sets, including the Markowitz (1952) and Roy feasible sets as special cases.

Specifically, Markowitz (1956) allowed the portfolio analyst to designate none, some, or all variables to be subject to nonnegativity constraints (as in Markowitz 1952) and the remaining variables to not be thus constrained (as in Roy). In addition to (or instead of) the budget constraint, the portfolio analyst could specify zero, one, or more linear equality constraints (sums or weighted sums of variables required to equal some constant) and/or linear inequality constraints (sums or weighted sums of variables required to be no greater or no less than some constant). A portfolio analyst can set down a system of constraints of these kinds such that no portfolio can meet all constraints. In this case, we say that the model is "infeasible." Otherwise, it is a "feasible model."

6

In addition to permitting any system of constraints, Markowitz (1956) made an assumption[2] that assured that if the model was feasible, then (as in Markowitz 1952) there was a unique feasible portfolio that minimized variance among portfolios with any given feasible E.

Markowitz (1956) showed that the set of efficient portfolios is piecewise linear in the general model, as in the special case of Markowitz (1952). Depending on the constraints imposed by the portfolio analyst, one of the linear pieces of the efficient set could extend without end in the direction of increasing E, as in the case of the Roy model. (Note that if the analysis contains 1,000 securities, the lines we are discussing here are straight lines in 1,000-dimensional "portfolio space." These lines may be hard to visualize and impossible to draw, but they are not hard to work with algebraically.)

Markowitz (1956) presented a computing procedure, the "critical line algorithm," that computes each corner portfolio in turn and the efficient line segment between them, perhaps ending with an efficient line "segment" on which feasible E increases without end. The formulas for the efficient line segments are all of the same pattern. Along a given "critical line," some of the variables that are required to be nonnegative are said to be OUT and are set to zero; the others are said to be IN and are free to vary. Variables not constrained to be nonnegative are always IN. On the critical line, some inequalities are called SLACK and are ignored; the others are BINDING and are treated (in the formula for the particular critical line) as if they were equalities. With its particular combination of BINDING constraints and IN variables, the formula for the critical line is the same as if the problem were to minimize V for various E subject to only equality constraints. In effect, OUT variables and SLACK constraints are deleted from the problem.

At each step, the algorithm uses the formula for the current critical line for easy determination of the next corner portfolio. The next critical line, which the current critical line meets at the corner, has the same IN variables and BINDING constraints as the current line except for *one* of the following—one variable moves from OUT to IN or moves from IN to OUT or one constraint moves from BINDING to SLACK or from SLACK to BINDING. This similarity between successive critical lines greatly facilitates the solution of one line when given the solution of the preceding critical line.[3]

Merton (1972) said, "The characteristics of the mean–variance efficient portfolio frontier have been discussed at length in the literature. However,

for more than three assets, the general approach has been to display qualitative results in terms of graphs" (p. 1851). I assume that at the time, Merton had not read Markowitz (1956) or Appendix A of Markowitz (1959).

Markowitz Portfolio Theory circa 1959

Markowitz (1959) was primarily written during the 1955–56 academic year while I was at the Cowles Foundation for Research in Economics at Yale at the invitation of Tobin. At the time, Tobin was already working on what was to become Tobin (1958), which is discussed in the next section.

I had left the University of Chicago for the RAND Corporation in 1951; my coursework was finished, but my dissertation (on portfolio theory) was still to be written. My RAND work had nothing to do with portfolio theory. So, my stay at the Cowles Foundation on leave from RAND provided an extended period when I could work exclusively on, as well as write about, portfolio theory. The following subsections summarize the principal ways in which my views on portfolio theory evolved during this period, as expressed in Markowitz (1959).

A Still More General Mean–Variance Analysis. The central focus of Markowitz (1959) was to explain portfolio theory to a reader who lacked advanced mathematics. The first four chapters introduced and illustrated mean–variance analysis, defined the concepts of mean, variance, and covariance, and derived the formulas for the mean and variance of a portfolio. Chapter 7 defined mean–variance efficiency and presented a geometric analysis of efficient sets, much like Markowitz (1952) but without the two errors noted previously. Chapter 8 introduced the reader to some matrix notation and illustrated the critical line algorithm in terms of a numerical example.

The proof that the critical line algorithm produces the desired result was presented in Appendix A of Markowitz (1959). Here, the result was more general than that in Markowitz (1956). The result in Markowitz (1956) made an assumption sufficient to assure that a unique feasible portfolio would minimize variance for any given E. Markowitz (1959) made no such assumption; rather, it demonstrated that the critical line algorithm will work for *any* covariance matrix. The reason it works is as follows: Recall that the equations for a critical line depend on which variables are IN and which are OUT. Appendix A showed that each IN set encountered in tracing out the

efficient frontier is such that the associated equations for the critical line are solvable.[4]

Models of Covariance. Markowitz (1959, pp. 96–101) argued that analysis of a large portfolio consisting of many different assets has too many covariances for a security analysis team to carefully consider them individually, but such a team can carefully consider and estimate the parameters of a model of covariance. This point was illustrated in terms of what is now called a single-index or one-factor (linear) model. The 1959 discussion briefly noted the possibility of a more complex model—perhaps involving multiple indexes, nonlinear relationships, or distributions that vary through time.

Markowitz (1959) presented no empirical analysis of the ability of particular models to represent the real covariance matrix (as in Sharpe 1963, Cohen and Pogue 1967, Elton and Gruber 1973, or Rosenberg 1974), and I did not yet realize how a (linear) factor model could be used to simplify the computation of critical lines, as would be done in Sharpe (1963) and in Cohen and Pogue.

The Law of the Average Covariance. Chapter 5 of Markowitz (1959) considered, among other things, what happens to the variance of an equally weighted portfolio as the number of investments increases. It showed that the existence of correlated returns has major implications for the efficacy of diversification. With uncorrelated returns, portfolio risk approaches zero as diversification increases. With correlated returns, even with unlimited diversification, risk can remain substantial. Specifically, as the number of stocks increases, the variance of an equally weighted portfolio approaches the "average covariance" (i.e., portfolio variance approaches the number you get by adding up all covariances and then dividing by the number of them). I now refer to this as the "law of the average covariance."

For example, if all securities had the same variance V_s and every pair of securities (other than the security with itself) had the same correlation coefficient ρ, the average covariance would be ρV_s and portfolio variance would approach ρV_s; therefore, portfolio standard deviation would approach $\sqrt{\rho}\sigma_s$. If the correlation coefficient that all pairs shared was, for example, 0.25, then the standard deviation of the portfolio would approach 0.5 times the standard deviation of a single security. In this case, investing in an unlimited number of securities would result in a portfolio whose standard deviation was 50 percent as great as that of a completely undiversified portfolio. Clearly, there is a qualitative difference in the efficacy of diversification depending on whether one assumes correlated or uncorrelated returns.

Semideviation. Semivariance is defined like variance (as an expected squared deviation from something) except that it counts only deviations below some value. This value may be the mean of the distribution or some fixed value, such as zero return. Semideviation is the square root of semivariance. Chapter 9 of Markowitz (1959) defined semivariance and presented a three-security geometric analysis illustrating how the critical line algorithm can be modified to trace out mean–semideviation-efficient sets. Appendix A presented the formal description of this modification for any number of securities and a proof that it works.

Mean and Variance of What? Why Mean and Variance? The basic ideas of Markowitz (1952) came to me sometime in 1950 while I was reading Williams (1938) in the Business School library at the University of Chicago. I was considering applying mathematical or econometric techniques to the stock market for my Ph.D. dissertation for the Economics Department. I had not taken any finance courses, nor did I own any securities, but I had recently read Graham and Dodd (1934), had examined Wiesenberger (circa 1950), and was now reading Williams.

Williams asserted that the value of a stock is the expected present value of its future dividends. My thought process went as follows: If an investor is only interested in some kind of expected value for securities, he/she must be only interested in that expected value for the portfolio, but the maximization of an expected value of a portfolio (subject to a budget constraint in nonnegative investments) does not imply the desirability of diversification. Diversification makes sense as well as being common practice. What was missing from the analysis, I thought, was a measure of risk. Standard deviation or variance came to mind. On examining the formula for the variance of a weighted sum of random variables (found in Uspensky 1937 on the library shelf), I was elated to see the way covariances entered. Clearly, effective diversification required avoiding securities with high covariance. Dealing with two quantities—mean and variance—and being an economics student, I naturally drew a trade-off curve. Being, more specifically, a student of T.C. Koopmans (see Koopmans 1951), I labeled dominated EV combinations "inefficient" and undominated ones "efficient."

The Markowitz (1952) position on the ques-

©Association for Investment Management and Research

tions used as the heading for this subsection differed little from my initial thoughts while reading Williams. Markowitz (1952) started by rejecting the rule that the "investor does (or should) maximize the discounted . . . [expected] value of future returns," both as a hypothesis about actual behavior and as a maxim for recommended behavior, because it "fails to imply diversification no matter how the anticipated returns are formed." Before presenting the mean–variance rule, Markowitz (1952) said:

> It will be convenient at this point to consider a static model. Instead of speaking of the time series of returns on the *i*th security ($r_{i,1}, r_{i,2}, \ldots,$ $r_{i,t}, \ldots$) we will speak of "the flow of returns" (r_i) from the *i*th security. The flow of returns from the portfolio as a whole is $R = \sum X_i r_i$.
> (pp. 45–46)

The flow of returns concept is not heard from after this point. Shortly, Markowitz (1952) introduced "elementary concepts and results of mathematical statistics," including the mean and variance of a sum of random variables. "The return (R) on the portfolio as a whole is a weighted sum of random variables (where the investor can choose the weights)." From this point forward, Markowitz (1952) was primarily concerned with how to choose the weights X_i so that portfolios will be mean–variance efficient.

Markowitz (1952) stated that its "chief limitations" are that "(1) we do not derive our results analytically for the *n*-security case; . . . (2) we assume static probability beliefs." This work expressed the intention of removing these limitations in the future. Markowitz (1956) and Appendix A of Markowitz (1959) addressed the first issue, and Chapter 13 of Markowitz (1959) addressed the second issue.

Chapters 10–12 of Markowitz (1959) reviewed the theory of rational decision making under risk and uncertainty. Chapter 10 was concerned with rational decision making in a single period with known odds; Chapter 11 reviewed many-period optimizing behavior (again, with known odds); Chapter 12 considered single- or many-period rational behavior when the odds might be unknown. The introduction in Chapter 10 emphasized that the theory reviewed there applies to an idealized rational decision maker with limited information but unlimited computing powers and is not necessarily a hypothesis about actual human behavior. This position contrasts with Markowitz (1952), which offered the mean–variance rule both as a hypothesis about actual behavior and as a maxim for recommended behavior.

Chapter 13 applied the theory of rational behavior—which was developed by John von Neumann and Oskar Morgenstern (1944), Leonard J. Savage (1954), Richard Bellman (1957), and others, and was reviewed in Chapters 10 through 12—to the problem of how to invest. It began with a many-period consumption–investment game and made enough assumptions to assure that the dynamic programming solution to the game as a whole would consist of maximizing a sequence of single-period "derived" utility functions that depended only on end-of-period wealth. Chapter 13 then asked whether knowledge of the mean and variance of a return distribution allows one to estimate fairly well the distribution's expected utility. The analysis here did *not* assume either normally distributed returns or a quadratic utility function (as in Tobin 1958). It did consider the robustness of quadratic approximations to utility functions. In other words, if you know the mean and variance of a distribution, can you approximate its expected utility? See also Levy and Markowitz (1979). Furthermore, Chapter 13 considered what kinds of approximations to expected utility are implied by other measures of risk.

The last six pages of the chapter sketched how one could or might ("could" in the easy cases, "might" in the hard cases) incorporate into a formal portfolio analysis considerations such as (1) consumer durables, (2) nonportfolio sources of income, (3) changing probability distributions, (4) illiquidities, and (5) taxes. As compared with later analyses, the Chapter 13 consumption–investment game was in discrete time rather than continuous time (as in Merton 1969), did not reflect the discovery of myopic utility functions (as did Mossin 1968 and Samuelson 1969), and did not consider the behavior of a market populated by consumers/investors playing this game. Its objective was to provide a theoretical foundation for portfolio analysis as a practical way to approximately maximize the derived utility function of a rational investor.

Tobin (1958)

Tobin was concerned with the demand for money as distinguished from other "monetary assets." Monetary assets, including cash, were defined by Tobin as "marketable, fixed in money value, free of default risk." Tobin stated:

> Liquidity preference theory takes as given the choices determining how much wealth is to be invested in monetary assets and concerns itself with the allocation of these amounts among cash and alternative monetary assets. (p. 66)

Tobin assumed that the investor seeks a mean–variance-efficient combination of monetary assets.

He justified the use of expected return and standard deviation as criteria on either of two bases: Utility functions are quadratic, or probability distributions are from some two-parameter family of return distributions.

Much of Tobin's article analyzed the demand for money when "consols"[5] are the only other monetary asset. The next-to-last section of the article was on "multiple alternatives to cash." Here, Tobin presented his seminal result now known as the Tobin Separation Theorem. Tobin assumed a portfolio selection model with n risky assets and one riskless asset, cash. Because these assets were monetary assets, the risk was market risk, not default risk. Holdings had to be nonnegative. Borrowing was not permitted. Implicitly, Tobin assumed that the covariance matrix for risky assets is nonsingular (or he could have made the slightly more general assumption of Markowitz 1956). Tobin showed that these premises imply that for a given set of means, variances, and covariances among efficient portfolios containing any cash at all, the proportions among risky stocks are always the same:

> . . . the proportionate composition of the non-cash assets is independent of their aggregate share of the investment balance. This fact makes it possible to describe the investor's decisions as if there were a single non-cash asset, a composite formed by combining the multitude of actual non-cash assets in fixed proportions. (p. 84)

The primary purpose of Tobin's analysis was to provide an improved theory of the holding of cash. He concluded that the preceding analysis

> . . . is a logically more satisfactory foundation for liquidity preference than the Keynesian theory. . . . Moreover, it has the empirical advantage of explaining diversification—the same individual holds both cash and "consols"— while the Keynesian theory implies that each investor will hold only one asset. (p. 85)

At a meeting with Tobin in attendance, I once referred to his 1958 article as the first capital asset pricing model (CAPM). Tobin declined the honor. It is beyond the scope of this article, which has a 1960 cutoff, to detail the contributions of William Sharpe (1964), John Lintner (1965), Jan Mossin (1966), and others in the development of capital asset pricing models. A comparison of the assumptions and conclusions of Tobin with those of Sharpe may, however, help locate Tobin in the development of today's financial economics.

Tobin contrasted his interest to mine as follows:

> Markowitz's main interest is prescription of rules of rational behavior for investors: the main concern of this paper is the implications

> for economic theory, mainly comparative statics, that can be derived from assuming that investors do in fact follow such rules. (p. 85, Note 1)

To this extent, at least, the focus of Sharpe (1964) is the same as that of Tobin. Tobin and Sharpe are also similar in postulating a model with n risky and one riskless security. The principal differences between the two are (1) a difference in assumption between their mathematical models and (2) the economic phenomena concerning which the respective models are asserted.

As for assumptions, Tobin assumed that one can invest (i.e., lend) at the risk-free rate. Sharpe assumed that the investor can either borrow or lend at the same rate. (Tobin usually assumed that the rate is zero, but he noted that this assumption is not essential.) This, perhaps seemingly small, difference between the two models makes for a substantial difference in their conclusions. First, if investors can borrow or lend all they want at the risk-free rate (and the covariance matrix among the n risky stocks is nonsingular), then *all* efficient portfolios consist of one particular combination of risky assets, perhaps plus borrowing or lending. The implication is that, in equilibrium, the market portfolio (plus borrowing or lending) is the *only* efficient portfolio. In the Tobin model, in contrast, if investors have heterogeneous risk tolerances—so some hold cash and others do not—the market portfolio can be quite inefficient, even when all investors have the same beliefs and all hold mean–variance-efficient portfolios (see Markowitz 1987, Chapter 12).

Probably the most remarkable conclusion Sharpe drew from his premises was that in CAPM equilibrium, the expected return of each security is linearly related to its beta and only its beta. This conclusion is not necessarily true in the Tobin model (see Markowitz 1987, Chapter 12).

The second major difference between the two works is that Sharpe postulated that his model applied to all securities, indeed all "capital assets," whereas Tobin postulated only that his model applied to "monetary assets." In fact, Tobin expressed doubts that cash should be considered risk free:

> It is for this reason that the present analysis has been deliberately limited . . . to choices among monetary assets. Among these assets cash is relatively riskless, even though in the wider context of portfolio selection, the risk of changes in purchasing power, which all monetary assets share, may be relevant to many investors.

Between them, Tobin's assumptions were more cautious; Sharpe's revolutionized financial economics.

Hicks (1935, 1962)

The Hicks (1962) article on liquidity included the following paragraph:

> It would obviously be convenient if we could take just one measure of "certainty"; the measure which would suggest itself, when thinking on these lines, is the standard deviation. The chooser would then be supposed to be making his choice between different total outcomes on the basis of mean value (or "expectation") and standard deviation only. A quite simple theory can be built up on that basis, and it yields few conclusions that do not make perfectly good sense. It may indeed be regarded as a straightforward generalisation of Keynesian Liquidity Preference. We would be interpreting Liquidity Preference as a willingness to sacrifice something in terms of mean value in order to diminish the expected variance (of the whole portfolio). Instead of looking simply at the single choice between money and bonds, we could introduce many sorts of securities and show the distribution between them determined on the same principle. It all works out very nicely, being indeed no more than a formalisation of an approach with which economists have been familiar since 1936 (or perhaps I may say 1935). [A footnote to the last sentence of this paragraph explained as follows:] Referring to my article, "A Suggestion for Simplifying the Theory of Money." *Economica* (February 1935). (p. 792)

The formalization was spelled out in a mathematical appendix to Hicks (1962) titled "The Pure Theory of Portfolio Investment" and in a footnote on p. 796 that presents an $E\sigma$ – efficient set diagram.

The appendix presented a mathematical model that is almost exactly the Tobin model with no reference to Tobin. The difference between the Hicks and Tobin models is that Hicks assumed that all correlations are zero whereas Tobin permitted any nonsingular covariance matrix. Specifically, Hicks presented the general formula for portfolio variance written in terms of correlations, rather than covariances, and then stated:

> It can, I believe, be shown that the main properties which I hope to demonstrate, remain valid whatever the r's; but I shall not attempt to offer a general proof in this place. I shall simplify by assuming that the prospects of the various investments are uncorrelated ($r_{jk} = 0$ when $k \neq j$): an assumption with which, in any case, it is natural to begin.

In the discussion that followed, Hicks (1962) derived the Tobin conclusion that among portfolios that include cash, there is a linear relationship between portfolio mean and standard deviation and that the proportions among risky assets remain constant along this linear portion of the efficient

frontier. In other words, Hicks presented what we call the Tobin Separation Theorem.

Hicks also analyzed the efficient frontier beyond the point where the holding of cash goes to zero. In particular, he noted that as we go out along the frontier in the direction of increasing risk and return, securities leave the efficient portfolio and do not return. (This last point is not necessarily true with correlated returns.[6])

Returning to the portion of the frontier that contains cash, if the Hicks (1962) results are, in fact, a formalization of those in Hicks (1935)—in the sense of transcribing into mathematics results that were previously described verbally—then the Tobin Separation Theorem should properly be called the Hicks or Hicks–Tobin Separation Theorem. Let us examine Hicks (1935) to see if it did anticipate Tobin as described in the appendix to Hicks (1962).

Within Hicks (1935), the topic of Section V is closest to that of "The Pure Theory of Portfolio Investment" in the appendix of Hicks (1962). Preceding sections of Hicks (1935) discussed, among other things, the need for an improved theory of money and the desirability of building a theory of money along the same lines as the existing theory of value. They also discussed, among other things, the relationship between the Hicks (1935) analysis and that of Keynes as well as the existence of "frictions," such as "the cost of transferring assets from one form to another." In Section IV, Hicks (1935) introduced risk into his analysis. Specifically, he noted, "The risk-factor comes into our problem in two ways: First, as affecting the expected period of investment, and second, as affecting the expected net yield of investment" (p. 7). In a statement applicable to both sources of risk, Hicks continued:

> Where risk is present, the *particular* expectation of a riskless situation is replaced by a band of possibilities, each of which is considered more or less probable. It is convenient to represent these probabilities to oneself, in statistical fashion, by a mean value, and some appropriate measure of dispersion. (No single measure will be wholly satisfactory, but here this difficulty may be overlooked.) (p. 8)

Hicks (1935) never designated standard deviation or any other specific measure as the measure he meant when speaking of risk. After discussing uncertainty of the period of the investment, he concluded Section IV thus:

> To turn now to the other uncertainty—uncertainty of the yield of investment. Here again we have a penumbra.... Indeed, without assuming this to be the normal case, it would be impossible to explain some of the most obvious of the observed facts of the capital market. (p. 8)

The theory of investment that Hicks (1935) presented in Section V may be summarized as follows:

> It is one of the peculiarities of risk that the total risk incurred when more than one risky investment is undertaken does not bear any simple relation to the risk involved in each of the particular investments taken separately. In most cases, the "law of large numbers" comes into play (quite how, cannot be discussed here). . . .
>
> Now, in a world where cost of investment was negligible, everyone would be able to take considerable advantage of this sort of risk reduction. By dividing up his capital into small portions, and spreading his risks, he would be able to insure himself against any large total risk on the whole amount. But in actuality, the cost of investment, making it definitely unprofitable to invest less than a certain minimum amount in any particular direction, closes the possibility of risk reduction along these lines to all those who do not possess the command over considerable quantities of capital. . . .
>
> By investing only a proportion of total assets in risky enterprises, and investing the remainder in ways which are considered more safe, it will be possible for the individual to adjust his whole risk situation to that which he most prefers, more closely than he could do by investing in any single enterprise. (pp. 9–10)

Hicks (1935) was a forerunner of Tobin in seeking to explain the demand for money as a consequence of the investor's desire for low risk as well as high return. Beyond that, there is little similarity between the two authors. Hicks (1935), unlike Tobin or the appendix to Hicks (1962), did not designate standard deviation or any other specific measure of dispersion as representing risk for the analysis; therefore, he could not show a formula relating risk on the portfolio to risk on individual assets. Hicks (1935) did not distinguish between efficient and inefficient portfolios, contained no drawing of an efficient frontier, and had no hint of any kind of theorem to the effect that all efficient portfolios that include cash have the same proportions among risky assets.

Thus, there is no reason why the theorem that currently bears Tobin's name should include any other name.

Marschak (1938)

Kenneth Arrow (1991) said of Marschak (1938):

> Jacob Marschak . . . made some efforts to construct an ordinal theory of choice under uncertainty. He assumed a preference ordering in the space of parameters of probability distributions—in the simplest case, the space of the mean and the variance. . . . From this for-

mulation to the analysis of portfolio selection in general is the shortest of steps, but one not taken by Marschak. (p. 14)

G.M. Constantinides and A.G. Malliaris (1995) described the role of Marschak (1938) as follows.

> The asset allocation decision was not adequately addressed by neoclassical economists. . . . The methodology of deterministic calculus is adequate for the decision of maximizing a consumer's utility subject to a budget constraint. Portfolio selection involves making a decision under uncertainty. The probabilistic notions of expected return and risk become very important. Neoclassical economists did not have such a methodology available to them. . . . An early and important attempt to do that was made by Marschak (1938) who expressed preferences for investments by indifference curves in the mean–variance space. (pp. 1–2)

An account of Marschak is, therefore, mandatory in a history of portfolio theory through 1960, if for no other reason than that these scholars judged it to be important. On the other hand, I know of one authority who apparently did not think the article to be important for the development of portfolio theory. My thesis supervisor was Marschak himself, and he never mentioned Marschak (1938). When I expressed interest in applying mathematical or econometric techniques to the stock market, Marschak told me of Alfred Cowles own interest in financial applications, resulting, for example, in Cowles 1939 work.[7] Then, Marschak sent me to Marshall Ketchum in the Business School at the University of Chicago for a reading list in finance. This list included Williams (1938) and, as I described, led to the day in the library when my version of portfolio theory was born. Marschak kept track of my work, read my dissertation, but never mentioned his 1938 article.

So, which authority is correct concerning the place of Marschak in the development of portfolio theory? Like Hicks, Marschak sought to achieve a better theory of money by integrating it with the General Theory of Prices. In the introductory section of the article, Marschak explained that

> to treat monetary problems and indeed, more generally, problems of investment with the tools of a properly generalized Economic Theory . . . requires, first, an extension of the concept of human *tastes*: by taking into account not only men's aversion for waiting but also their desire for safety, and other traits of behaviour not present in the world of perfect certainty as postulated in the classical static economics. Second, the *production conditions*, assumed hereto to be objectively given, become, more realistically, mere subjective expectations of the investors—and *all* individ-

©Association for Investment Management and Research

uals are investors (in any but a timeless economy) just as all market transactions are investments. The problem is: to explain the objective quantities of goods and claims held at any point of time, and the objective market prices at which they are exchanged, given the subjective tastes and expectations of the individuals at this point of time. (p. 312)

In the next five sections, Marschak presented the equations of the standard economic analysis of production, consumption, and price formation. Section 7 dealt with choice when outcomes are random. No new equations were introduced in this section. Rather, Marschak used the prior equations with new meanings:

We may, then, use the previous formal setup if we reinterpret the notation: $x, y \ldots$ shall mean, not future yields, but parameters (e.g., moments and joint moments) of the joint-frequency distribution of future yields. Thus, x may be interpreted as the mathematical expectation of first year's meat consumption, y may be its standard deviation, z may be the correlation coefficient between meat and salt consumption in a given year, t may be the third moment of milk consumption in second year, etc. (p. 320)

Marschak noted that people usually like high mean and low standard deviation; also, "they like meat consumption to be accompanied by salt consumption" (i.e., z as well as x in the preceding quotation "are positive utilities" as opposed to standard deviation, y, which is "a disutility"). He noted that people "like 'long odds' (i.e., high positive skewness of yields." However, it "is sufficiently realistic . . . to confine ourselves, for each yield, to two parameters only: the mathematical expectation ('lucrativity') and the coefficient of variation ('risk')."

So, is Marschak's article a forerunner of portfolio theory or not? Yes and no. It is not a step (say, beyond Hicks 1935) toward *portfolio* theory because it does not consider *portfolios*. The means, standard deviations, and correlations of the analysis, including the means (and so on) of end products consumed, appear directly in the utility and transformation functions with no analysis of how they combine to form moments of the investor's portfolio as a whole. On the other hand, Marschak's 1938 work is a landmark on the road to a theory of markets whose participants act under risk and uncertainty, as later developed in Tobin and the CAPMs. It is the farthest advance of economics under risk and uncertainty prior to the publication of von Neumann and Morgenstern (1944).

Williams (1938)

The episode reported previously in which I discovered the rudiments of portfolio theory while reading Williams occurred in my reading early parts of the book. Later in the book, Williams observed that the future dividends of a stock or the interest and principal of a bond may be uncertain. He said that, in this case, probabilities should be assigned to various possible values of the security and the mean of these values used as *the* value of the security. Finally, he assured readers that by investing in sufficiently many securities, risk can be virtually eliminated. In particular, in the section titled "Uncertainly and the Premium for Risk" (starting on p. 67 in the chapter on "Evaluation by the Rule of Present Worth"), he used as an example an investor appraising a risky 20-year bond "bearing a 4 per cent coupon and selling at 40 to yield 12 per cent to maturity, even though the pure interest seems to be only 4 per cent." His remarks apply to any investor who "cannot tell for sure" what the present worth is of the dividends or of the interest and principal to be received:

Whenever the value of a security is uncertain and has to be expressed in terms of probability, the correct value to choose is the mean value. . . . The customary way to find the value of a risky security has always been to add a "premium for risk" to the pure interest rate, and then use the sum as the interest rate for discounting future receipts. In the case of the bond under discussion, which at 40 would yield 12 per cent to maturity, the "premium for risk" is 8 per cent when the pure interest rate is 4 per cent.

Strictly speaking, however, there is no risk in buying the bond in question if its price is right. Given adequate diversification, gains on such purchases will offset losses, and a return at the pure interest rate will be obtained. Thus the *net risk* turns out to be nil. To say that a "premium for risk" is needed is really an elliptical way of saying that payment of the full face value of interest and principal is not to be expected on the average.

In my 1952 article, I said that Williams's prescription has the investor

diversify his funds among all those securities which give maximum expected return. The law of large numbers will insure that the actual yield of the portfolio will be almost the same as the expected yield. This rule is a special case of the expected returns–variance of returns rule. . . . It assumes that there is a portfolio which gives both maximum expected return and minimum variance, and it commends this portfolio to the investor.

This presumption, that the law of large numbers applies to a portfolio of securities, cannot be accepted. The returns from securities are too intercorrelated. Diversification cannot eliminate all variance.

That is still my view. It should be noted, however, that Williams's "dividend discount model" remains one of the standard ways to estimate the security means needed for a mean–variance analyses (see Farrell 1985).

Leavens (1945)

Lawrence Klein called my attention to an article on the diversification of investments by Leavens, a former member of the Cowles Commission. Leavens (1945) said:

> An examination of some fifty books and articles on investment that have appeared during the last quarter of a century shows that most of them refer to the desirability of diversification. The majority, however, discuss it in general terms and do not clearly indicate why it is desirable.

Leavens illustrated the benefits of diversification on the assumption that risks are independent. However, the last paragraph of Leavens cautioned:

> The assumption, mentioned earlier, that each security is acted upon by independent causes, is important, although it cannot always be fully met in practice. Diversification among companies in one industry cannot protect against unfavorable factors that may affect the whole industry; additional diversification among industries is needed for that purpose. Nor can diversification among industries protect against cyclical factors that may depress all industries at the same time.

Thus, Leavens understood intuitively, as did Shakespeare 350 years earlier, that some kind of model of covariance is at work and that it is relevant to the investment process. But he did not incorporate it into his formal analysis.

Leavens did not provide us with his reading list of "some fifty books and articles." This omission is fortunate because I am probably not prepared to read them all and the reader is surely not ready to read accounts of them. Let us assume, until some more conscientious student of this literature informs us otherwise, that Leavens was correct that the majority discussed diversification in general terms and did "not clearly indicate why it is desirable." Let us further assume that the financial analysts who did indicate why it is desirable did not include covariance in their formal analyses and had not developed the notion of an efficient frontier. Thus, we conclude our survey with Leavens as representative of finance theory's analysis of risk as of 1945 and, presumably, until Roy and Markowitz in the 1950s.

The End of the Beginning

One day in 1960, having said what I had to say about portfolio theory in my 1959 book, I was sitting in my office at the RAND Corporation in Santa Monica, California, working on something quite different, when a young man presented himself at my door, introduced himself as Bill Sharpe, and said that he also was employed at RAND and was working toward a Ph.D. degree at UCLA. He was looking for a thesis topic. His professor, Fred Weston, had reminded Sharpe of my 1952 article, which they had covered in class, and suggested that he ask me for suggestions on a thesis topic. We talked about the need for models of covariance. This conversation started Sharpe out on the first of his (ultimately many) lines of research, which resulted in Sharpe (1963).

For all we know, the day Sharpe introduced himself to me at RAND could have been exactly 10 years after the day I read Williams. On that day in 1960, there was *no* talk about the possibility of using portfolio theory to revolutionize the theory of financial markets, as done in Sharpe (1964), nor was there any inkling of the flood of discoveries and applications, many by Sharpe himself, that were to occur in investment theory and financial economics during the next four decades.

©**Association for Investment Management and Research**

Notes

1. Given the assumptions of Markowitz (1952), if more than one portfolio has maximum feasible E, only one of these portfolios will be efficient, namely, the one with the smallest V. This one will be reached by the "tracing out" process described.

2. The assumption was that V is strictly convex over the set of feasible portfolios. This assumption is weaker than requiring the covariance matrix to be nonsingular.

3. In the text, I am discussing the shape of efficient sets in portfolio space. As observed in Markowitz (1952), the set of efficient EV combinations is piecewise parabolic, with each line segment in portfolio space corresponding to a parabolic segment in EV space. As discussed previously, Markowitz (1956) understood that successive parabolas meet in such a way that efficient V as a function of E is strictly convex. Markowitz (1956) noted that typically there is no kink where two successive efficient parabola segments meet: The slope of the one parabola equals that of the other at the corner portfolio where they meet. Markowitz (1956) did, however, note the possibility of a kink in the efficient EV set if a certain condition occurred, but the 1956 work did not provide a numerical example of a problem containing such a kink. For numerical examples of problems with kinks in the efficient EV set, see Dybvig (1984) and Chapter 10 of Markowitz (1987).

4. The equations in Markowitz (1956) also depended on which inequalities were BINDING. Markowitz (1959) wrote inequalities as equalities, without loss of generality, by introducing "slack variables" as in linear programming. The critical line algorithm works even if the constraint matrix, **A**, as well as the covariance matrix, **C**, is rank deficient. The critical line algorithm begins with George Dantzig's (1963) simplex algorithm to maximize E or determine that E is unbounded. The simplex algorithm introduces "dummy slacks," some of which remain in the critical line algorithm if **A** is rank deficient (see Markowitz 1987, Chapters 8 and 9). Historically, not only did I have great teachers at the University of Chicago, including Jacob Marschak, T.C. Koopmans, Milton Friedman, and L.J. Savage, but I was especially fortunate to have Dantzig as a mentor when I worked at RAND.

5. Government bonds in Great Britain, originally issued in 1751, that (similarly to an annuity) pay perpetual interest and have no date of maturity.

6. See Chapter 11 in Markowitz (1987) for a three-security example of risky assets in which an asset leaves and later reenters the efficient portfolio. Add cash to the analysis in such a way that the coming and going of the security happens above the tangency of the line from $(0, r_0)$ to the frontier. Perhaps, if you wish, add a constant to all expected returns, including r_0, to assure that $r_0 \geq 0$.

7. The Cowles Commission for Research in Economics, endowed by Alfred Cowles, was affiliated with the University of Chicago at the time. Marschak was formerly its director. I was a student member.

References

Arrow, Kenneth. 1991. "Cowles in the History of Economic Thought." *Cowles Fiftieth Anniversary*. The Cowles Foundation for Research in Economics at Yale University:1–24.

Bellman, Richard. 1957. *Dynamic Programming*. Princeton, NJ: Princeton University Press.

Cohen, K.J., and J.A. Pogue. 1967. "An Empirical Evaluation of Alternative Portfolio Selection Models." *Journal of Business*, vol. 40, no. 2 (April):166–193.

Constantinides, G.M., and A.G. Malliaris. 1995. In *Handbooks in OR & MS*. Edited by R. Jarrow, et al. Vol. 9. Amsterdam, Netherlands: Elsevier Science B.V.

Cowles, A., 3rd, and Associates. 1938. *Common-Stock Indexes, 1871–1937*. Bloomington, IA: Principia Press.

Dantzig, G.B. 1963. *Linear Programming and Extensions*. Princeton, NJ: Princeton University Press.
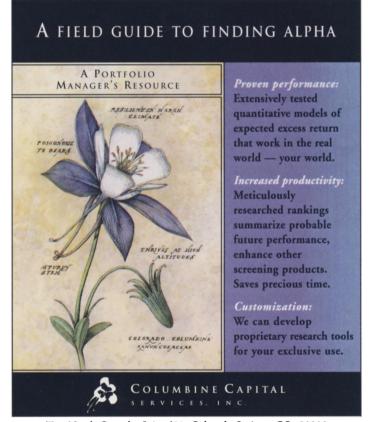
Dybvig, Philip H. 1984. "Short Sales Restrictions and Kinks of the Mean Variance Frontier." *Journal of Finance*, vol. 39, no. 1 (March):239–244.

Elton, E.J., and M.J. Gruber. 1973. "Estimating the Dependence Structure of Share Prices." *Journal of Finance*, vol. 28, no. 5 (December):1203–32.

Farrell, J.L., Jr. 1985. "The Dividend Discount Model: A Primer." *Financial Analysts Journal*, vol. 41, no. 6 (November/December):16–19, 22–25.

Graham, Benjamin, and David L. Dodd. 1934. *Security Analysis*. New York: McGraw-Hill.

Hicks, J.R. 1935. "A Suggestion for Simplifying the Theory of Money." *Economica* (February):1–19.

———. 1962. "Liquidity." *Economic Journal*, vol. 72 (December):787–802.

Koopmans, T.C. 1951. "Analysis of Production as an Efficient Combination of Activities." In *Activity of Production and Allocation*, 7th ed. Edited by T.C. Koopmans. 1971. New Haven, CT: Yale University Press.

Leavens, D.H. 1945. "Diversification of Investments." *Trusts and Estates*, vol. 80 (May):469–473.

Levy, H., and H.M. Markowitz. 1979. "Approximating Expected Utility by a Function of Mean and Variance." *American Economic Review*, vol. 69, no. 3 (June):308–317.

Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics*, vol. 47, no. 1 (February):13–37.

Markowitz, Harry M. 1952. "Portfolio Selection." *Journal of Finance*, vol. 7, no. 1 (March):77–91.

———. 1956. "The Optimization of a Quadratic Function Subject to Linear Constraints." *Naval Research Logistics Quarterly*, vol. 3:111–133.

———. 1959. *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons.

———. 1987. *Mean–Variance Analysis in Portfolio Choice and Capital Markets*. Oxford, U.K.: Basil Blackwell.

Marschak, J. 1938. "Money and the Theory of Assets." *Econometrica*, vol. 6:311–325.

Merton, R.C. 1969. "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case." *Review of Economics and Statistics*, vol. 51, no. 3 (August):247–259.

———. 1972. "An Analytic Derivation of the Efficient Portfolio Frontier." *Journal of Financial and Quantitative Analysis*, vol. 7, no. 4 (September):1851–72.

Mossin, J. 1966. "Equilibrium in a Capital Asset Market." *Econometrica*, vol. 35, no. 4 (October):768–783.

———. 1968. "Optimal Multiperiod Portfolio Policies." *Journal of Business*, vol. 41, no. 2 (April):215–229.

Rosenberg, Barr. 1974. "Extra-Market Components of Covariance in Security Returns." *Journal of Financial and Quantitative Analysis*, vol. 9, no. 2 (March):263–273.

Roy, A.D. 1952. "Safety First and the Holding of Assets." *Econometrica*, vol. 20, no. 3 (July):431–449.

Samuelson, P.A. 1969. "Lifetime Portfolio Selection by Dynamic Stochastic Programming." *Review of Economics and Statistics*, vol. 51, no. 3 (August):239–246.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: John Wiley & Sons

Sharpe, William F. 1963. "A Simplified Model for Portfolio Analysis." *Management Science*, vol. 9, no. 2 (January):277–293.

———. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance*, vol. 19, no. 3 (September):425–442.

Tobin, James. 1958. "Liquidity Preference as Behavior towards Risk." *Review of Economic Studies*, vol. 25, no. 1 (February):65–86.

Uspensky, J.V. 1937. *Introduction to Mathematical Probability*. New York: McGraw-Hill.

von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. 3rd ed. 1967. Princeton, NJ: Princeton University Press.

Wiesenberger, A., and Company. *Investment Companies*. New York, annual editions since 1941.

Williams, J.B. 1938. *The Theory of Investment Value*. Cambridge, MA: Harvard University Press.