Portfolio Theory: As I Still See It

Harry M. Markowitz

Harry Markowitz Company, San Diego, California 92109;
email: HarryHMM@aol.com

Annu. Rev. Financ. Econ. 2010. 2:1–23

First published online as a Review in Advance on September 29, 2010

The *Annual Review of Financial Economics* is online at financial.annualreviews.org

This article's doi:
10.1146/annurev-financial-011110-134602

1941-1367/10/1205-0001$20.00

Key Words

MPT, Friedman-Savage, prospect theory, stochastic dominance, capital asset pricing model

Abstract

This essay summarizes my views on (*a*) the foundations of portfolio theory and its applications to current issues, such as the choice of criteria for practical risk-return analysis, and whether some form of risk-return analysis should be used in fact; (*b*) hypotheses about actual financial behavior, as opposed to idealized rational behavior, including two proofs of the fact that expected-utility maximizers would never prefer a multiple-prize lottery to all single-prize lotteries, as asserted in one of my 1952 papers; and (*c*) a simple proof of the theorem (which was initially greeted with some skepticism, especially by referees) that investors in capital asset pricing models do not get paid for bearing risk.

INTRODUCTION

I thank the editorial committee members of this volume, in particular the editors Robert Merton and Andrew Lo, for this opportunity to express my views on topics that concern me most regarding financial theory and practice, especially as they relate to my own contributions to this field. This article is divided into three sections, each with various subsections. The first section deals with my fundamental assumptions concerning the practical use of mean-variance analysis. I never—at any time!—assumed that return distributions are Gaussian. To the contrary, on the one occasion that a colleague and I investigated the form of the return-generating distribution we concluded (in Markowitz & Usmen 1996a, 1996b) that, among the very broad class of distributions in the Pearson Family, daily log returns on the S&P 500 were most likely generated by a student-*t* distribution with between four and five degrees of freedom.

Nor did I ever assume that the investor's utility function is quadratic. Rather, I noted that quadratic approximations to traditional utility functions are often quite good over a surprisingly large range of returns. For example, in the case of the logarithmic utility function, the approximation is quite close from a 30% or 40% loss to a 40% or 50% gain. Therefore, if the return on the portfolio-as-a-whole is usually within this range, and never "too" far beyond it, then expected log utility will be closely approximated by a function of mean and variance—regardless of the shape of the return distribution.

Along these lines, this paper begins by recapitulating the Markowitz (1959) analysis of the utility functions implicit in the use of alternate measures of risk. The analysis is extended to include VaR and CVaR, not covered in my 1959 text. I also present here my current view as to what criteria to use—for central tendency as well as risk—when return distributions are too spread out for mean-variance to apply.

The next section of this paper is concerned with a hypothesis proposed by Markowitz (1952b) to explain the existence of both the buying of lottery tickets and the buying of insurance, sometimes by the same economic agent. This hypothesis avoids some questionable implications of the hypothesis by Friedman & Savage (1948), which also sought to explain the existence of insurance and lotteries. My utility function noted in Markowitz (1952b) is similar to the value function in Kahneman & Tversky (1979) in that it has an inflection point at or near current wealth, but it differs in that my function is concave below and convex above the inflection point, whereas theirs is convex below and concave above. I cite arguments pro and con for these two hypotheses, including Levy & Levy's (2004) ingenious examples of pairs of probability distributions *P* and *Q*, such that *P* PT (prospect theory) dominates *Q*, and *Q* Markowitz (1952b) dominates *P*.

The topic that interests me most in this section concerns the assertion (without proof) in the last footnote of Markowitz (1952b) that an expected-utility maximizer would never prefer a multiple-prize lottery to all single-prize lotteries. I present here two proofs for this assertion. The first, which I had in mind when I wrote Markowitz (1952b), assumes that there are a finite number of possible dollar outcomes of the lottery, and the lottery purchaser is free to choose any probabilities of these outcomes, subject to two constraints: (*a*) that the chosen values be nonnegative numbers that sum to one and (*b*) that the lottery be "fair," i.e., have a zero expected return or, more generally, have a specified expected return (equal to the lottery vendor's profit). I then give a more general proof without the finiteness requirement.

Then I show that one cannot conclude from the widespread existence of multiple-prize lotteries that buyers are not expected-utility maximizers. If potential ticket buyers have diverse utility functions, and if the lottery vendor can hold only one lottery at a time—e.g., one Irish sweepstakes or one New Jersey lottery—then it may maximize the vendor's profit to offer a multiple-prize lottery. Finally, assuming that further research will show that lottery purchasers do in fact prefer multiple-prize lotteries, I propose a nonexpected-utility-maximizing hypothesis that describes the phenomenon, albeit without providing a psychological explanation for it.

The final section of this paper is concerned with the capital asset pricing model (CAPM). It reiterates a complaint of mine that if

$$\sum X_i = 1$$

is the only constraint on the choice of portfolio, then negative X_i are not an accurate—in fact, they are an absurd—representation of real-world short positions. However, if the above equation is not assumed as the only constraint, nor is it assumed that all investors can borrow all they want at the risk-free rate, then the market is typically not an efficient portfolio. Also, a linear relationship is not typically found between expected returns and betas. Thus, the many hundreds of hours spent by our colleagues trying to establish whether a linear relationship exists between expected return and beta, in effect, seeks to determine empirically whether, e.g., the reader can borrow the same amount as can the U.S. Treasury and for the same interest rate. In conclusion, I provide a simple proof of the fact that—if one makes all the assumptions of either the Sharpe-Lintner or the "Roy" CAPM and therefore derives a linear relationship between expected returns and betas—this linear relationship does not imply that CAPM investors get paid for bearing risk.

FUNDAMENTAL ASSUMPTIONS OF PORTFOLIO THEORY AND THEIR APPLICATION TO ISSUES OF THE DAY

In this section, I review my fundamental assumptions concerning portfolio theory, particularly as they relate to issues such as whether—and when—mean-variance analysis is applicable. Also discussed is what should be used in its stead when such analysis is not applicable.

Markowitz Versus Markowitz

My views concerning the fundamental assumptions of portfolio theory were formed while writing Markowitz (1959). Without major exception, this is still how I see these matters. These views are to be contrasted with those presented in Markowitz (1952a). Differences between Markowitz (1952a) and Markowitz (1959) include the following:

1. Markowitz (1952a) proposes mean-variance both in a positive sense (as a hypothesis about investor behavior) and in a normative sense (as a recommendation for portfolio selection). Markowitz (1959) makes no mention of mean-variance as a positive hypothesis about investor behavior. This was not an oversight. At the time Markowitz (1959) was written, I was not prepared either to defend or to renounce mean-variance as a hypothesis.
2. Markowitz (1952a) presents no rationale for the normative use of mean-variance. It simply proposes it. In contrast, Markowitz (1959) argues that under certain conditions a carefully chosen portfolio from a mean-variance efficient frontier will approximately maximize the investor's expected utility. I review this argument below.

3. The analyses of Markowitz (1952a) are quite different from those of Markowitz (1959) concerning the relationship between the single-period mean-variance analysis and the many-period "game" in which it is embedded. On the one hand, Markowitz (1952a) offers two such analyses. The first considers the investor as seeking high mean and low variance of the present value of future returns. The other assumes that available opportunities (with respect to means, variances, and covariances) are in equilibrium: The same opportunities present themselves each period. Markowitz (1959), on the other hand, assumes that the investor seeks to maximize expected value for the "game-as-a-whole." Therefore, the single-period expected utility function that mean-variance seeks to approximately maximize is the dynamic programming "derived" utility function.[1]

Portfolio Theory and the Theory of Rational Behavior Under Uncertainty

The fundamental assumptions of Markowitz (1959) are presented late in the book because I feared that practitioners would not read a book that began with axioms of choice under uncertainty (see part IV, ch. 12 and 13 in particular). Following Savage (1954), Markowitz (1959) proposes axioms concerning rational decision making under uncertainty. These imply that the rational decision maker maximizes expected utility using personal probability beliefs where objective probabilities are not known. The axioms presented in Markowitz (1959, ch. 12) are simpler than those provided in Savage (1954), in part, because I assume that there are only a finite number of possible hypotheses that might be true—perhaps in the billions or trillions—and only a finite number of possible outcomes that can occur from any experiment or lifetime.

Another reason that the Markowitz (1959) model is simpler is because Savage (1954) excludes any consideration of objective probability, whereas Markowitz (1959) assumes that the decision maker can rank alternatives that involve the possibility of objective probabilities. This does not mean that objective probabilities do in fact exist. Rather, it only implies that hypotheses that involve the assumption of objective probability can be stated. An example would be the hypothesis that the logarithm of the ratio of some market index on day $t+1$ to that on day t is normally distributed.

Mean-Variance Approximations to EU

Markowitz (1959) asserts that if $U_t(R_t)$ can be approximated closely enough by a quadratic for a sufficiently wide range of returns, then $E(U)$ will be approximately equal to some function, $f(E,V)$, of mean and variance. In particular, I suggested two methods of approximating a given $U(R)$ by a quadratic:

$$Q_Z(R) \cong U(0) + U'(0)R + 0.5U''(0)R^2 \tag{1}$$

and

$$Q_E(R) \cong U(E) + U'(E)(R-E) + 0.5U''(E)(R-E)^2, \tag{2}$$

where $\cong$ indicates is approximately equal to and a prime denotes differentiation.

[1]The consumption/investment game of Markowitz (1959, ch. 13) is a discrete-time dynamic game, not the now widely used continuous-time game introduced by Robert Merton. See Merton (1990).

The first of these is a Taylor expansion around $R = 0$; the second is around $R = E$. The two functions of E and V, obtained by taking the expected values of these, are

$$f_Z(E,V) = E[Q_Z(r)] = U(0) + U'(0)E + 0.5U''(0)(E^2 + V) \tag{3}$$

and

$$f_E(E,V) = E[Q_E(R)] = U(E) + 0.5U''(E)V. \tag{4}$$

For example, when $U = Ln(1 + R)$, then f_Z and f_E are

$$f_Z^L(E,V) = E - (E^2 + V)/2 \tag{5}$$

and

$$f_Z^L(E,V) = Ln(1 + E) - V/[2(1 + E)^2]. \tag{6}$$

As shown in **Table 1** (whose first three columns are the contents of table 2 of Markowitz 1959) for returns between a 30% loss and a 40% gain on the portfolio-as-a-whole there is little difference between $Ln(1 + R)$ and its Q_Z. Therefore, for probability distributions confined to this range, EU and $f(E,Z) = EQ_Z$ must be close—regardless of the shape of the distribution. As we leave this range, utility U and the quadratic approximation Q_Z begin to separate noticeably, though not too badly even in the range [−0.4, 0.5]. As the range of possible returns increases further, however, the approximation deteriorates at an increasing rate. In particular, $Ln(1 + R) \rightarrow -\infty$ as $R \rightarrow -1$, whereas $Q_Z \rightarrow -1.5$. As R increases, Q_Z reaches a maximum at $R = 1$ and then heads in the wrong direction.

Thus, for choice among return distributions that are mostly within the [−0.3, 0.4] range—on the portfolio-as-a-whole—and do not fall outside this range "too far, too often," the $E[Ln(1 + R)]$ maximizer will almost maximize expected utility by an appropriate choice from the mean-variance efficient frontier. Markowitz (1959, ch. 6) examines this

Table 1 Comparison of $Ln(1 + R)$ with $R - (½)R^2$ and $R - (½)(R^-)^2$

R	$Ln(1 + R)$	$R - (½)(R)^2$	$R - (½)(R^-)^2$
−.50	−.69	−.63	−.63
−.40	−.51	−.48	−.48
−.30	−.36	−.35	−.35
−.20	−.22	−.22	−.22
−.10	−.11	−.11	−.11
.00	.00	.00	.00
.10	.10	.10	.10
.20	.18	.18	.20
.30	.26	.26	.30
.40	.34	.32	.40
.50	.41	.38	.50

approximation using the annual returns on the nine securities used in the illustrative portfolio analysis discussed in chapter 2. Young & Trent (1969) investigate approximations to $E[Ln(1 + R)]$ or, equivalently, the geometric mean (GM) and find the mean-variance approximation to be quite robust. Levy & Markowitz (1979) do similar experiments for a variety of utility functions and historical returns on mutual fund portfolios. Their general conclusion is that, typically, for probability distributions of returns such as those observed on mutual fund portfolios, if the investor chooses carefully from among mean-variance efficient portfolios, the investor will approximately maximize expected utility, even if the investor does not know its utility function explicitly. Dexter et al. (1980), Pulley (1981, 1983), Kroll et al. (1984), Simaan (1987), Ederington (1995), and Hlawitschka (1994) find similar conclusions for various kinds of situations and utility functions. But Grauer (1986) reminds us that the mean-variance approximation will break down if too much leverage is used, as in CAPM examples without nonnegativity constraints.

Alternate Approaches to Portfolio Selection

As explained above, the Markowitz (1959) defense of mean-variance is not unconditional. Rather, it asserts the existence of situations in which returns on the portfolio-as-whole are mostly confined to a range in which the investor's utility function can be approximated sufficiently well by a quadratic, and occasional departures from this range are "not too serious." As Levy & Markowitz (1979) and others confirm, for many utility functions and for distributions of returns, such as historical returns of investment companies, functions of mean-variance supply robust approximations to expected utility.

The question remains: Which portfolio selection process should be used if mean-variance is not applicable? Major alternatives include the following:

1. Use other measures of risk or return in a risk-return analysis
2. Determine the investor's utility function explicitly and maximize its expected value
3. Do not optimize; instead, use constraints and guidelines
4. Proceed intuitively

I briefly comment on the other three alternatives before focusing on the first. In particular, I examine the utility functions implicit in the use of alternatives examined in Markowitz (1959), plus alternatives not known in 1959, namely VaR and CVaR.

If executed properly, the best alternative is the second one. The danger is that an inappropriate utility function will be chosen because of analytic or computational convenience. Furthermore, a single utility function does not reflect the fact that the investor's utility function may, appropriately, change over time as a function of circumstances, perhaps including the investor's age and wealth. Risk-return analysis finesses the problem of determining the investor's authentic, current utility function by involving the investor (or his or her investment advisor) in trade-off judgments. The problem with taking the human out of the loop is that, not only may the assumed utility function make the wrong trade-offs, but the process may also be too much of a black box for the investor or his or her advisor to understand what is going on and to override it when appropriate.

The constraints and guidelines mentioned in the third alternative are routinely implemented as constraints in a mean-variance analysis. This raises the question of whether it is better to use such constraints on their own or as part of a risk-return analysis. The answer is that rarely, if ever, do the intuitively desired constraints completely specify one and only one portfolio. (If the portfolio analyst tried to do so, he or she would probably overspecify the portfolio and end up with a system with no feasible solution.) So then, how is the analyst to choose among the many portfolios that meet the constraints? (For a discussion of an intuition-only decision, see the comments below concerning the fourth alternative.) If the choice is made using some kind of decision rule (such as, "Try to get as close to equal weighting as possible, subject to the constraints"), either the procedure employed would be "closet" risk-return (e.g., equal weighting subject to constraints is the same as mean-variance with all assets having the same means, variances, and covariances), or it is probably a procedure unrelated to the maximization of expected utility. In the latter case, I can only echo the reasons I gave in 1959 based on arguments by Von Neumann & Morgenstern (1944) and Marschak (1950) as to why I accept the maximization of expected utility as essential to rational action in the face of risk and uncertainty.

Concerning the fourth alternative to proceed intuitively, the behavioral finance literature is full of ways that seat-of-the-pants investing leads to irrational choices. For example, Barber & Odean (2000) show that investors who fancy that they can outperform the market by active trading typically make a lot of money for their brokers and little, if any, for themselves.

During a conference discussion session, Peter Bernstein described the haphazard way in which institutional portfolios were formed in the 1950s and earlier. He concluded with the statement, "Now you have a process." My thought at the time was, "Now I understand what I started." Of course, "the process" is more than mean-variance analysis á la Markowitz (1959). It often uses a top-down approach pioneered by Brinson and colleagues (see Brinson et al. 1986, 1991; Ibbotson & Kaplan 2000; Statman 2000). It may use a factor model of either expected returns or covariance (see Sharpe 1963, Rosenberg 1974, Fama & French 1992). It often makes use of data such as that of Morningstar (2010), which continues that of Ibbotson and of Ibbotson and Sinquefield. The process is no longer confined to large institutional investors, but it is now available to individual investors via software their investment advisors can use or via 401(k) advisory services such as William Sharpe's Financial Engines or Guided Choice (to which I consult).

Judgment plays an essential role in the proper application of risk-return analysis for individual and institutional portfolios. For example, the estimates of mean, variance, and covariance of a mean-variance analysis should be forward-looking rather than purely historical. Even a purely historical approach to parameter estimation involves judgment as to what estimation period to use. The choice of constraints for a risk-return analysis also involves judgment. In particular, "[c]onstraints are added (in part, at least) because the investor seeks protection against contingencies whose probability of 'disutility' is underrated by the mean-variance approximation or, possibly, by the parameter estimation procedure" (my foreword to Jacobs & Levy 2000).

In general, Markowitz (1959) and my current view is that constraints and judgment should play an essential role in portfolio choice. However, the implications for the portfolio-as-a-whole of such constraints and judgments should be thought through, ideally with an expected-utility analysis with a properly chosen utility function. In the absence of the latter, I believe that some form of risk-return analysis is the best alternative.

Alternate Measures of Risk

Beginning with the "Utility Functions and Measures of Risk" section, Markowitz (1959, p. 286) considers the utility functions implicitly associated with various risk measures. The applicable basic theorem states the following:

> If a decision maker maximizes expected utility, and its action depends only on a function $g(Er, Ef(R))$, of ER and $Ef(R)$, then the utility function, whose expectation the decision maker maximizes, is

$$U = a + bR + cf(R). \tag{7}$$

Thus, the linearity of the expected-utility rule implies the linearity of the g function. In particular, if action in accord with expected utility depends only on ER and ER^2, therefore only on ER and $V(R) = ER^2 - (ER)^2$, then U must be a quadratic. As previously discussed, even though this is an unattractive approximation as $R \to \infty$, because U reaches a maximum and then heads in the wrong direction, it is a sufficient approximation for many concave utility functions and for return distributions such as historical returns on investment company portfolios. The questions I explore now are, What are the approximating utility functions implicit in the use of other risk measures, and are any of these more attractive than mean-variance when the latter is, or is not, appropriate?

Semivariance

Markowitz (1959, ch. 9) defines the semivariance around the constant b, S_b, to be

$$S_b = E((R - b)^-)^2, \tag{8}$$

where X^- represents the negative part of X. Thus the use of mean and semivariance as portfolio selection criteria implicitly approximates the investor's utility function by

$$U = a + cR + d[(R - b)^-]^2. \tag{9}$$

Markowitz (1959) offers various pros and cons for the use of mean-variance versus mean-semivariance. I now believe that the comparison can be sharpened with the aid of **Table 1**. As in Markowitz (1959, p. 121), the first three columns compare log utility with a quadratic approximation centered around $R = 0$. Here I have added a fourth column showing the implied utility function, Equation 9, for semivariance with $b = 0$, namely S_0, expected squared loss. Specifically, the fourth column shows

$$U = R - \tfrac{1}{2}(R^-)^2. \tag{10}$$

For $R \leq 0$, the mean-variance and mean-S_0 approximations are identical, namely loss squared. Therefore any difference in the ability of the respective functions to approximate $U = Ln(1 + R)$ is for $R > 0$. However, the table shows that for $R \in [0.00, 0.50]$, the quadratic is superior to the squared-loss approximation of Equation 10.

Thus, to approximate expected $U = Ln(1 + R)$ for return distributions with negligible probability of return exceeding 50% on the portfolio-as-a-whole, mean-variance is superior to mean-semivariance. **Table 1** applies to the logarithmic utility function, but similar conclusions—in favor of mean-variance over mean-semivariance—presumably apply for the broad range of utility functions for which Levy, Markowitz, and others find mean-variance to be a satisfactory approximation. The empirical results show that a quadratic $Q(R)$ with

$$\begin{aligned} Q(R) &= U(R) \\ Q'(R) &= U'(R) \\ Q''(R) &= U''(R) \end{aligned}$$

at some central point, such as $R = 0$ or $R = E$, is an effective approximation to expected utility. The problem with the approximation $S_0(R)$ in Equation 9 is that $S''(R) = 0$ for R $> b$.

The comparison in **Table 1** uses S_b with $b = 0$. A more favorable comparison is obtained using a larger b. For example, $R - \frac{1}{2}R^2$ reaches a maximum at $R = 1.0$. The utility function implicit with the use of mean-$S_{1.0}$ as criteria is identical to that of mean-variance for $R \leq 1.0$, and then it is horizontal rather than decreasing. With this choice of b, mean-S_b has a more desirable implicit utility function than mean-variance. However, $S_{1.0}$ can hardly be called downside risk. It may, instead, be called a "don't penalize large gains the way mean-variance does" criterion. The discussion below shows that S_b for b near 0 (such as S_0 in particular) is quite attractive when combined with a different measure of "return."

Other Measures of Risk

In addition to variance and semivariance, Markowitz (1959, ch. 13) examines the following measures of risk:

- Expected loss
- Expected absolute deviation
- Probability of loss
- Maximum loss

The chapter finds all these measures less satisfactory than either variance or semivariance. In particular, the utility function associated with use of mean and expected loss is

$$U = a + bR + cR^{-}. \tag{11}$$

In terms of **Table 1**, for $R > 0$ the utility function suffers from the same problem as the mean-S_0 approximation, i.e., it is linear for such returns. In addition, it suffers from the more serious drawback that it is linear for $R < 0$. Among distributions with the same probability distribution of gains, it is risk-neutral with respect to the probability distribution of losses.

A similar objection applies to the use of expected value and the mean absolute deviation from $R = 0$. In this case, the implicit utility function is

$$U = a + bR + c|R|. \tag{12}$$

Both Equations 11 and 12 consist of two linear pieces, as do

$$U = a + dR + c(R - b)^{-} \tag{13}$$

and

$$U = a + dR + c|R - b|. \tag{14}$$

The difference between the utility functions in Equations 11 and 12 versus those in Equations 13 and 14 is that the kink where the two linear pieces meet is at $R = b$ for the latter two functions rather than specifically at $R = 0$ as in the former. In none of these cases does the approximation show diminishing marginal utility as R increases, given that $R > b$ or $R < b$.

Although Markowitz (1959) does not demonstrate such, it can be shown that there is no $f(R)$, such that expected $f(R)$ equals the mean absolute deviation from the mean[2]:

$$MAD = E|R - ER|.$$

Markowitz (1959) does, however, show that there is no function $f(r)$ such that $Ef(r)$ equals maximum loss.

VaR and CVaR

Neither VaR nor CVaR is the expected value of some function $f(R)$. To see this, consider a situation in which the decision maker must choose among probability distributions $(p_1, p_2, \ldots, p_n)$ of dollar outcomes $d_1 < d_2 \ldots < d_n$. Suppose, for concreteness, that

$$\sum_{i=1}^{4} p_i = .04$$

and consider VaR and CVaR at the 0.05 level as a function of

$$p_5 = .01 + \delta$$

for δ in a small interval $(-\varepsilon,\varepsilon)$. For $\delta < 0$, VaR is d_6 and CVaR is expected loss given $R \leq d_6$. At $\delta = 0$, VaR jumps to d_5, and CVaR jumps to the expected loss given $R \leq d_5$.

Expected utility is usually justified by a set of axioms, such as the three axioms of Markowitz (1959, ch. 10). One of the Markowitz axioms says that choice is, in a certain sense, a continuous function of probabilities $p_1, \ldots, p_n$.[3] CVaR can also be derived from basic assumptions about desirable risk measures, namely that the risk measure be coherent in the sense provided by Pflug (2000); see also Rockafellar & Uryasev (2000). Rather than a general comparison of sets of assumptions that justify expected utility versus those that justify VaR or CVaR, the question here is specifically whether preference among alternatives should change radically with an infinitesimal change in probabilities. My own view remains, as in Markowitz (1959), that continuity of preference is desirable. More generally, I still accept the expected-utility maximum. I therefore count the discontinuous nature of VaR and CVaR as a "black mark" against these proposed measures of risk.

[2]That there is no $f(R)$ such that $Ef(R) = E|R - ER|$ may be illustrated with the binomial distribution with $\mathrm{Prob}(R = 1) = p$ and $\mathrm{Prob}(R = 0) = 1 - p$. If there were a and b such that $f(1) = a$, $f(0) = b$, and $Ef(R) = E|R - ER|$, then, for every p, we would have $pa + (1 - p)b \equiv p|1 - p| + (1 - p)|-p|$. For $p\varepsilon(0, 1)$ this requires $(a - b)p + b = 2(p - p^2)$. But the expression on the left is linear and the expression on the right is quadratic with a unique maximum at $p = \frac{1}{2}$.

[3]Specifically, Axiom II in chapter 10 assumes (concerning preferences among probability distribution P, Q and R) that if P is preferred to Q and Q is preferred to R, then there is a probability p of obtaining P versus $(1 - p)$ of obtaining R that is exactly as good (preference-wise) as having Q, i.e., $P \,\mathrm{pref}\, Q \,\mathrm{pref}\, R$ implies that there exists a p such that $pP + (1 - p)Q \approx R$.

It is true that, if an investor maximizes the expected value of a utility function, a small change in probabilities can lead to a large shift in optimum portfolio. However, the new optimum will have almost the same expected utility, measured with either the old or the new probability distribution. For example, suppose that an investor sought to maximize the expected value of a linear function

$$U = a + bER \qquad b > 0.$$

Then expected utility is maximized by maximizing ER. If two securities have almost the same ER, then 100% of the optimum portfolio will be invested in the security with the slightly greater ER. A small shift in probabilities can lead max ER to shift to the other security, and the optimum portfolio will shift from 100% of the one to 100% of the other. But the utility of the utility-maximizing portfolio will change little. In the case of a ranking function of the form $g(ER, VaR)$ or $g(ER, CVaR)$, not only can the optimum portfolio jump from one end of portfolio space to the other, but the portfolio that was formerly optimum can now be deemed horrible in terms of the ranking function g.

CVaR also suffers from the following—at least equally serious—drawback: Consider the criterion

$$CLoss = E(R|R < L)$$

for some "large" loss $L < 0$. This is the conditional expected return given the loss. If L happens to equal VaR then $CLoss = CVaR$. $CLoss$ is an expected value

$$CLoss = E((R - |L|))^{-}.$$

Thus the implied utility function

$$R - b(R - |L|)^{-}$$

consists of two linear pieces, and it suffers from the same objections as applied to mean absolute deviation.

An Alternate Measure of "Return"

For the reasons stated, I do not consider any of the above alternatives to be a satisfactory answer to the question of what type of risk measure to use in a risk-return analysis if return distributions are too spread out for functions of mean and variance to approximate expected utility well. In particular, we saw that ES_b, mean-semivariance about a return $R = b$, has the problem that it is linear for $R \geq b$. In this range, it does not have diminishing marginal utility of wealth. For example, its use implies indifference between receiving \$(100,000,000 + b) with certainty versus a 50-50 chance of \$$b$ or \$(200,000,000 + b). Even for less extreme cases, the lack of risk aversion among returns greater than b seems undesirable. Conversely, for returns less than b, its implied approximating utility function is the same as that of mean-variance.

One cure for these problems is to combine the semideviation as a measure of risk with the geometric mean (GM) as the measure of return. Because

$$\log(1 + GM) = E \log(1 + R),$$

the utility function implicit in such an analysis would be

$$U = a + c\log(1+R) - d[(R-b)^{-}]^2. \tag{15}$$

This is concave (therefore, risk averse) and strictly increasing everywhere. It approaches $-\infty$ as $R \downarrow -1.0$. If $d = 0$ in Equation 15, then Equation 15 reduces to the logarithmic utility function. This is generally considered the utility function applicable for "long-run" investing (see Markowitz 2005). If $d > 0$ in Equation 15, then the investor would trade some return in the long run for greater stability in the short run. Conditional on return exceeding b, the $GM\text{-}S_b$ investor would act to maximize growth in the long run. When return below b is possible, the $GM\text{-}S_b$ investor chooses more cautiously.

The use of variance with GM is not as attractive as the use of S_b with GM because

$$U = a + b\log(1+R) - cR^2$$

for $c > 0$ reaches a maximizing R and turns downward for higher R.

It is much easier to compute a mean-variance efficient frontier than a $GM\text{-}S_b$ efficient frontier. So long as $E\log(1 + R)$ can be estimated sufficiently well from E and V, the economical way to generate a $GM\text{-}V$ frontier is to generate a mean-variance efficient frontier and then plot GM on the return axis. However, when distributions are too spread out for mean-variance approximations to be adequate, the extra expense is justified for deriving efficient portfolios on the $GM\text{-}S_b$ frontier. This expense is not only computational, but also includes additional estimation requirements because $E\log(1 + R)$ is not a function of first and second moments only. See Vander Weide (2010) for a further discussion of the use of GM in portfolio analysis.

THE UTILITY OF WEALTH

Friedman & Savage (1948; F-S) propose a utility function to explain the existence of both lotteries and insurance, including the existence of economic agents (e.g., individuals or families) who purchase both. The F-S function, illustrated in **Figure 1**, has utility on the vertical axis and wealth on the horizontal axis. It consists of two concave regions separated by a convex region. As in Markowitz (1952b), the figure here is drawn with a straight-line tangent to the curve at two points, c and d. Economic agents with less wealth than c are considered poor; those with wealth greater than d are considered rich.

Markowitz (1952b) uses this construction to demonstrate some questionable implications of the F-S hypothesis. For example, two "middle-class" agents, each with wealth halfway between c and d, would prefer a bet that would send one to c and the other to d over any other fair bet. More than that, they would prefer to take this bet rather than each remain at the intermediate wealth $(c + d)/2$. But middle-class economic agents do not typically engage in such wild gambles.

The F-S hypothesis also implies that poor people, with wealth below c, do not buy lottery tickets and that "almost rich" people, with wealth slightly below d, would not buy insurance to insure against wealth falling to c. On the contrary, the F-S hypothesis implies that they would prefer to engage in a fair gamble that, if won, would bring their wealth up to d and, if lost, would plunge it down to c. All these implications seem contrary to common sense and common observation.

Markowitz (1952b) concludes that the only point on the F-S utility function that has reasonable implications is the left-most inflection point. This has a concave portion to its left—i.e., the buying of insurance against loss—and a convex portion to its right—i.e., the

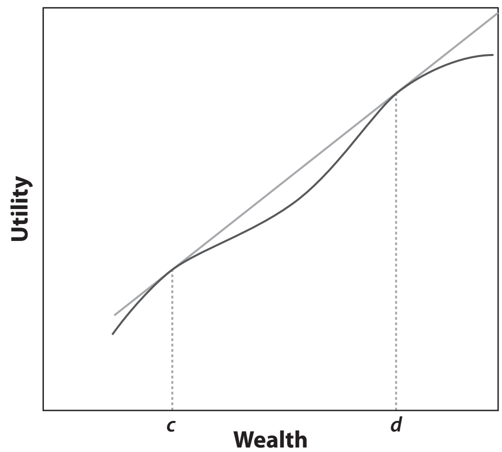

Figure 1

Friedman-Savage utility function with the Markowitz (1959) double tangency line.

buying of lottery tickets. Markowitz (1952b) hypothesizes that this inflection point was located at the agent's "customary wealth," which equaled current wealth except in the case of recent windfall gains or losses. In the case of a recent windfall gain, the agent's current wealth would move temporarily into the convex portion of the curve to the right of the inflection point. As a result, the agent would become more carefree in his or her risk taking. [Thaler & Johnson (1990) refer to this as gambling with the house's money.] Conversely, in case of a windfall loss, the agent would move temporarily into the concave region to the left of the inflection point and act more cautiously.

Using the St. Petersburg paradox and its generalization by Menger (1967) and then applying it to possible large losses as well as possible large gains, Markowitz (1952b) assumes that the hypothesized utility function is bounded above and below. This implies the existence of another convex region to the left of the concave region left of the "central" inflection point and yet another concave region to the right of the convex region right of the central inflection point (see **Figure 2**).

I have been asked how I reconcile the quite different decision rules of Markowitz (1952a) and Markowitz (1952b). It is easier to reconcile Markowitz (1952b) with Markowitz (1959), because Markowitz (1952a) proposed mean-variance efficiency both as a maxim and as a hypothesis, whereas Markowitz (1959) maintained mean-variance as a maxim but was mum about it as a hypothesis. Thus there is no inconsistency between recommending mean-variance efficiency to investors—especially fiduciaries who should not gamble with other people's money—as in Markowitz (1959), versus hypothesizing the Markowitz (1952b) utility function for the class of agents who both gamble and buy insurance.

I should also reconcile the assumption here that utility is (and, I believe, should be) bounded above and below with my recommendation of GM and S_b as criteria for portfolio selection with Equation 15 as their implied utility function. I believe that the boundedness assumption is correct, and the use of $\log(1 + R)$ is an approximation that is good enough for making choices among a restricted class of alternatives; just as the quadratic approximation implicit in the use of mean and variance was good enough for an even more restricted class of alternatives. For example, the utility function in Equation 15 would

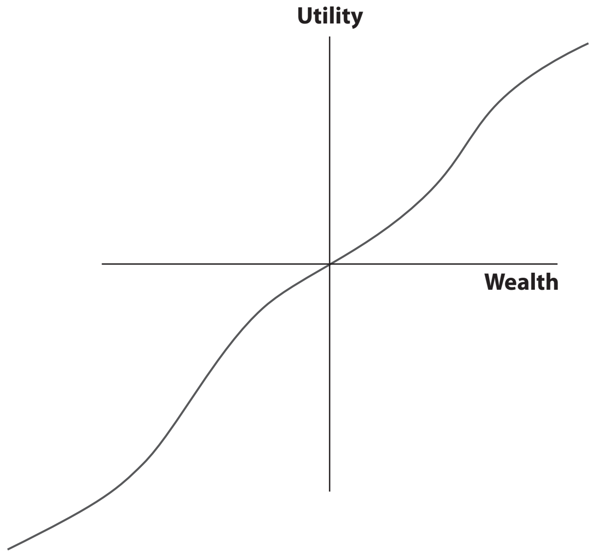

Figure 2

Markowitz (1952b) utility-of-wealth function.

prefer to lose 99% of a portfolio's value with certainty rather than have one chance in a million of losing 100% versus a 0.999999 chance of breaking even. But I imagine that most investors would prefer to take a miniscule chance of a 100% loss rather than settle for a definite 99% loss. Thus utility (bounded below in this instance) is correct, whereas Equation 15 is just good enough—usually.

Markowitz Versus Prospect Theory

Similar to the Markowitz (1952b) hypothesis, the prospect theory of Kahneman & Tversky (1979) measures utility (which they refer to as the "value") as a function of the deviation from current wealth, rather than as a function of wealth itself. The two main differences between Kahneman & Tversky (1979) and Markowitz (1952b) are as follows: (*a*) The Kahneman & Tversky (1979) utility function is convex to the left of the origin and concave to the right, whereas the Markowitz (1952b) utility function is concave to the left and convex to the right. (*b*) Markowitz (1952b) maximizes expected utility,

$$EU = \sum_{i=1}^{n} p_i u_i,$$

using probabilities $p_1 \ldots, p_n$, whereas Kahneman & Tversky (1979) use weights

$$\pi_i = \prod(p_i)$$

to maximize

$$\tilde{E}V = \sum_{i=1}^{n} \pi_i \mathrm{v}_i.$$

Once weight functions are introduced, the formula no longer represents an expected-utility hypothesis; that is, choice no longer maximizes a linear function of probabilities. Below, I present an alternate nonexpected-utility hypothesis and give my reasons for it there.

Concerning the Markowitz (1952b) hypothesis—concave to the left, convex to the right—versus the Kahneman & Tversky (1979) hypothesis—convex to the left, concave to the right—Levy & Levy (2004) define two new forms of stochastic dominance, analogous to first- and second-order stochastic dominance. Levy & Levy (2004) call one of them prospect theory dominance (PT-dominance) and the other Markowitz dominance (M-dominance). A probability distribution P is said to PT-dominate a distribution Q if P would be preferred to Q by all agents with the prospect theory value function of Kahneman & Tversky (1979). By contrast, P M-dominates Q if it would be preferred by all decision makers with a Markowitz (1952b) utility function. Ingeniously, Levy & Levy (2004) present distributions P and Q such that P PT-dominates Q, whereas Q M-dominates P. Thus, if we assume that agents use one or the other of these two utility functions, then their choice between P and Q reveals the shape of their utility/value function. The Levy & Levy (2004) survey of investor choices tends to favor M-dominance over PT-dominance, but this has been the subject of an ongoing controversy.

Kontek (2009) has an alternate view of the Markowitz (1952b) versus the PT hypotheses. The PT combination of weights and values are in response to a series of choices among gambles reported by Kahneman & Tversky (1979).

All the observations of Kahneman & Tversky (1979) are explained given the following:

1. One uses probabilities in calculating expected utilities rather than functions of probabilities $\pi(p)$.
2. One assumes that the utility function is essentially the Markowitz (1952b) utility function.
3. However, one must also let the scale of the horizontal axis in **Figure 2** depend on the choices involved. Equivalently, one may leave the horizontal scale the same, while spreading out or contracting the curve depending on available choices.

Kontek (2009) argues that such rescaling of the horizontal axis is analogous to (or perhaps a special case of) Weber's Law. This implies, for example, that in the presence of a very bright (or very noisy) room, a small difference in brightness or sound level would not be noticed, whereas such would be noticed in a dimly lit (or quiet) room.[4]

It does not violate the expected-utility hypothesis to assume that an agent's utility function changes with time. However, the Kontek hypothesis that, at a given time, the decision maker will use one or another utility function depending on the alternatives offered does violate the expected-utility hypothesis. There is much evidence to support the notion that human decision-making is not consistent with expected-utility maximization; this is most famously shown in Allais (1953) as well as Kahneman & Tversky (1979). The next section notes that a frequently observed market phenomenon also seems to contradict the expected-utility hypothesis. The subsequent section presents a formula for describing this apparently nonexpected-utility-maximizing behavior.

[4]Subsequently, Kahneman & Tversky (1979) modified their choice of weights to avoid a problem with stochastic dominance. The theory with the modified weights is referred to as cumulative prospect theory. Kontek's proposal is unaffected by comparing it with prospect theory versus cumulative prospect theory because his proposal uses the original, untransformed probabilities.

Multiple-Prize Lotteries

With few, if any, exceptions, lotteries have multiple prizes. If ticket holders do not win the big prize, they have many chances to win lesser prizes. Yet, in the last footnote, Markowitz (1952b) asserts the following:

> The existence of multiple lottery prizes with various sized prizes *may* contradict the theory presented. If we are forced to concede that the individual (lottery-ticket buyer) prefers, say, a fair multiple prize lottery to all other fair lotteries, then my hypothesis cannot explain this fact. . .Nor can any hypothesis which assumes that people maximize expected utility.

This footnote offers no proof of its assertion that if an expected-utility-maximizing agent could design its own lottery, subject to it being a fair game, the agent would never prefer a multiple-prize lottery to all others. I here give two proofs of this: The first, which I had in mind when I wrote the footnote, applies when the lottery is confined to a finite number of prespecified dollar outcomes. The second allows a larger class of probability distributions of lottery outcomes.

Theorem 1

Suppose that an agent can design its own "gamble" by choosing probabilities $p_1, p_2, \ldots, p_n$ that the gamble will have dollar outcomes $d_1, d_2, \ldots d_n$. His choice is subject to two constraints: (*a*) The p_i must be probabilities; therefore,

$$\sum_{i=1}^{n} p_i = 1 \tag{16a}$$

$$p_i \geq 0 \qquad i = 1, \ldots, n. \tag{16b}$$

(*b*) The game has a given expected loss or gain, i.e.,

$$\sum_{i=1}^{n} p_i d_i = k. \tag{17}$$

In particular, if $k = 0$, then the lottery is required to be a "fair" game. Define a "possible outcome" of the lottery as d_i with $p_i > 0$. If the decision maker maximizes expected utility

$$EU = \sum_{i=1}^{n} p_i u_i \tag{18}$$

for some given utility vector $u_i, \ldots, u_n$, then the agent will never prefer a gamble with three or more possible outcomes to all gambles with only one or two possible outcomes.

Proof. The maximization of the linear function, Equation 18, subject to the two linear constraints, Equations 16a and 17, in nonnegative variables, Constraints 16b, is a $2 \times n$ linear programming problem. Dantzig's (1963) simplex algorithm is based on a theorem concerning any $m \times n$ linear program to the effect that if an optimum solution exists then one exists with $X_i = 0$ for n - m variables. (One or more of the m "basis" variables of the solution may also be zero "accidentally.") For our $2 \times n$ linear program, the fact that the constraint set is closed and bounded implies that an optimum solution exists; therefore, an optimum solution exists with all but one or two of the p_i equal to zero.

QED. If only one p_i is not zero, then the two constraint equations imply that $p_i = 1$ for i with $d_i = k$. Otherwise we have positive p_i and p_j for i and j with $d_i < k$, $d_j > k$. In the case of a "lottery," d_i is the cost of the lottery ticket and d_j is its single payoff net of the cost of the ticket.

Therefore, if there were only a finite number of possible lottery prizes, the expected-utility maximizer would never prefer a multiple-prize lottery to all single-prize lotteries. The next theorem considers a decision maker who chooses a cost C of a lottery ticket, a probability p of winning some prize, and a probability distribution $P(R)$ of possible payoffs R if the ticket wins. C is chosen subject to a budget constraint $C \leq \bar{C}$. The expected-utility-maximizing agent is assumed to choose a probability distribution so as to maximize

$$EU = (1-p)U(-C) + pEU(R) \tag{19}$$

subject to

$$-(1-p)C + pE(R) = k. \tag{20}$$

As before, if $k = 0$, then the agent chooses the fair lottery that maximizes the agent's expected utility.

I make the following assumption concerning the agent's utility function. Consider all lotteries with cost $C \leq \bar{C}$, a single payoff R (net of cost), and probabilities of winning p such that

$$-(1-p)C + pR = k. \tag{21}$$

I assume that there is at least one among these single-prize lotteries that has greater expected utility

$$EU = (1-p)U(-C) + pU(R) \tag{22}$$

than any other such lottery. This rules out the possibility that, for a given C, EU in Equation 22 increases indefinitely as $R \to \infty$. By assumption, C is not allowed to increase without limit.

Theorem 2

If C_0, p_0, $P_0(R)$ maximize EU among single-prize lotteries, with $C \leq \bar{C}$, which satisfy Equation 21, then this single-prize lottery maximizes expected utility among C, p, $P(R)$ that satisfy Equation 20.

Proof. Let C, p, $P(R)$ be a cost of ticket, the probability of a win, and the probability distribution of the winning amount if the ticket wins, which satisfy Equation 20. EU in Equation 19 is most naturally thought of as the expected utility associated with a two-step process: First decide (with probability p) whether the ticket wins, and then draw the winning amount from the distribution $P(R)$. It can also be thought of as a different two-step process: First, draw R from $P(R)$ to decide what the payoff will be in case the ticket wins, and then decide (with probability p) whether the ticket wins. Given a particular $\tilde{R}$ drawn from $P(R)$, the conditional expected utility of the game, given this $\tilde{R}$, is

$$E(U|\tilde{R}) = (1-p)U(-C) + pU(\tilde{R}). \tag{23}$$

The expected utility of the game then is

$$EU = \underset{R}{E}(E(U|R)), \tag{24}$$

i.e., it is the expected value, over all R, of the conditional expected utility given R as shown in Equation 23. But

$$\begin{aligned} E(U|R) &= (1-p)U(-C) + pU(R) \\ &\leq (1-p_0)U(-C_0) + p_0U(R_0) \end{aligned}$$

for all R. Therefore,

$$EU = E(E(U|R)) \leq (1-p_0)U(-C_0) + p_0U(R_0).$$

QED. Thus, given our assumptions, the expected-utility maximizer never prefers a multiple-prize lottery to all single-prize lotteries with the same expected loss. However, this does not imply that the existence of multiple-prize lotteries contradicts the expected-utility hypothesis. It could be that offering a lottery with multiple prizes is the correct business decision in a market where expected-utility maximizers have varying utility functions.

For example, suppose that some state contemplates offering a lottery with cost C. In addition, it chooses one of the following:

- \$1000 prize with the probability of any one ticket winning equal to 0.001
- \$1,000,000 prize with a win probability per ticket of 0.000001 ($= 1/10^6$)
- Multiple-prize lottery with a payoff of $\frac{\text{payoff}}{\text{probability}}\left|\frac{1{,}000}{1/2{,}000}\right|\frac{1{,}000{,}000}{1/2{,}000{,}000}$

If $C = \$1.0$, then the lottery offers a fair bet. Presumably the state will set $C > \$1$.

Next, assume that potential buyers of lottery tickets are expected-utility maximizers with one or the other of two utility functions (**Table 2**). For simplicity we assume that a lottery purchaser will buy at most one ticket, either because it has a very negative value of $U(-2C)$ or because $2C$ exceeds its gambling budget constraint $\bar{C}$. The expected utilities of the three possible lotteries for the two different utility functions are presented in **Table 3**.

Because not buying the ticket has utility $U(0) = 1$, players of Type A will not buy a ticket for a lottery with the \$1,000,000 payoff only, and players of Type B will not buy a ticket for the lottery with the \$1,000 payoff only. However, both will buy a ticket for the multiple-prize lottery. Thus the state will have greater revenue with the multiple-prize lottery, even though the population consists of expected-utility maximizers.

Table 2 Utility functions for potential buyers of lottery tickets

Buyer type	$U(-C)$	$U(0)$	$U(1{,}000)$	$U(1{,}000{,}000)$
A	0	1	1010	999,000
B	0	1	999	1,010,000

Table 3 Expected utilities of three lotteries to two classes of players

	Payoffs		
Player type	\$1,000	\$1,000,000	Both
A	1.010	0.999	1.0045
B	0.999	1.010	1.0045

A Hypothesis

Thus it remains to be determined whether the existence of multiple-prize lotteries is the result of one of the following:

1. A preference for multiple-prize lotteries by ticket buyers who therefore cannot be expected-utility maximizers or
2. A heterogeneous population of expected-utility maximizers.

If the second scenario turns out to be the complete explanation, then no further hypotheses are needed. By constrast, if the first scenario is at least partly correct, then an additional hypothesis is needed to describe such preferences.

The one I offer here is not based on any psychological principles. It is just descriptive of the observed phenomena. If this description proves to be reasonably accurate, then psychologists and behavioral finance specialists can perhaps explain why.

Machina (1987) reconciles observed contradictions to the expected-utility hypothesis by postulating that action is a not-necessarily-linear function of probabilities, i.e., in the finite case with outcomes $d_1, d_2, \ldots d_n$, the agent maximizes a function

$$\phi = f(p_1, p_2, \ldots, p_n).$$

Machina assumes that the partial derivatives $\partial\phi/\partial p_i$ exist; therefore, f is locally linear. As a result, the expected-utility hypothesis applies locally.

But the observation that f is locally linear does not explain the preference for multiple-prize lotteries. This would be implied if we could adequately approximate f by a quadratic

$$Q = a + \sum b_i p_i + \sum\sum c_{ij} p_i p_j$$

with a positive definite (c_{ij}) matrix. It would be convenient if (c_{ij}) were diagonal, but that is not plausible. For example, suppose that a lottery offered prizes of \$1,000,000; \$1,001,000; \$1,002,000, etc. I doubt that an agent who prefers multiple-prize lotteries would prefer that distribution to a lottery with a \$1,000,000 grand prize and the remainder of the prize money spread among lesser amounts, such as \$50,000; \$10,000; \$1000; and even some \$50 or \$100 prizes with probabilities p_i inversely proportional to payoffs d_i and with expected outcome satisfying Equation 20. This suggests that c_{ij} decline with some measure of the distance between d_i and d_j.

CAPITAL ASSET PRICING MODEL

The CAPM of Sharpe (1964) and Lintner (1965) assumes that all investors seek mean-variance efficiency and that all have the same beliefs. Within this CAPM, the S-L assumption also applies.

S-L Assumption

The S-L assumption holds that all investors can lend all they have or borrow all they want at the risk-free rate, R_f. It concludes that (*a*) the market portfolio is a mean-variance efficient portfolio and that (*b*) the excess return of each security is proportional to its regression (β_i) against the market portfolio, i.e.,

$$e_i = k\beta_i \qquad i = 1, \ldots, n, \tag{25}$$

where $e_i = ER_i - R_f$. An alternate CAPM also assumes that investors seek mean-variance efficiency and have the same beliefs but, instead of making the S-L assumption, they make the Roy (1952) assumption, namely that

$$\sum X_i = 1, \tag{26}$$

where X_i is the fraction invested in security i.

Negative X_i are interpreted as short sales, but Equation 26 is not how short sales are actually constrained. For example, Equation 26 permits as feasible the portfolio (−1000, 1001, 0, . . . , 0). This would correspond to placing \$1000 with your broker, shorting \$1,000,000 worth of stock A, then using the proceeds from this sale plus your \$1000 to buy \$1,001,000 worth of stock B. If you are a CAPM aficionado and actually believe that Equation 26 is how short sales are constrained, ask your broker to explain "Reg T." Or see Jacobs et al. (2005) for a description and formalization of real-world short constraints.

When the Roy (1952) constraint is substituted for the S-L constraint, it still follows that the market is an efficient portfolio. Rather than Equation 25, CAPM with the Roy constraint implies a linear relationship between expected returns and betas:

$$E(R_i) = a + b\beta_i \qquad \text{for } i = 1, \ldots, n. \tag{27}$$

Equation 25 is a special case of Equation 27. Markowitz & Todd (2000) and Markowitz (2005) show that if we neither assume that the S-L assumption nor the Roy assumption hold, then it will typically not be true that the market is a mean-variance efficient portfolio, nor would there typically be—even approximately—a linear relationship between expected return and beta.

Thus the market can consist of economic units all of whom have the same correct beliefs and who seek and achieve mean-variance efficiency. Yet, the market portfolio is not a mean-variance efficient portfolio, provided that the participants can neither borrow all they want at the risk-free rate, nor can they short and use the proceeds to buy long. Markowitz (2005) also shows that, even if some investor were subject to either the S-L constraint or the Roy constraint and others were not, then the former would not arbitrage away the mean-variance inefficiency of the market portfolio. A corollary is that there is no such thing as a representative investor.

Behavioral finance economists, such as Barber & Odean (2000), provide ample evidence that the market contains inefficient investors. However, an observation that no empirical linear relationship exists between $E(R_i)$ and β_i is not additional evidence of such. It can be explained by the obvious fact—which can be confirmed by any banker or broker—that few if any investors can borrow all they want at the risk-free rate or use the proceeds of a short sale to buy long.

CAPM Investors Do Not Get Paid for Bearing Risk

Equations 25 and 27 are usually interpreted as showing that, given CAPM assumptions, the investor is paid to take on undiversifiable (market) risk. Markowitz (2008) shows that this is not a correct interpretation of these equations. The argument is quite simple in the case of the S-L CAPM when returns are uncorrelated. I review this case here and refer the reader to Markowitz (2008) for the analysis with a general covariance matrix.

The way Equation 25 arises is that, to minimize V for given E in the S-L CAPM world, each investor I must choose a portfolio $(X_1^I, \ldots, X_n^I)$ that satisfies

$$\sum_{j=1}^{n} X_i^I \sigma_{ij}, = k^I e_i \qquad \text{for } i = 1, \ldots, n. \tag{28}$$

This varies from one investor to another because of varying risk tolerance reflected in k^I.

Taking a sum over investors, weighted by investor wealth, we find that Equation 28 also holds for the market portfolio $(X_1^M, \ldots, X_n^M)$:

$$\sum_{j=1}^{n} X_i^M \sigma_{ij}, = k^M e_i \qquad i = 1, \ldots, n. \tag{29}$$

The left-hand side is the covariance between return R_i and the market portfolio R^M.

Thus, Equation 29 may be written

$$Cov(R_i, R^M) = k^M e_i \qquad i = 1, \ldots, n. \tag{30}$$

Dividing both sides by the variance of R^M gives us Equation 25, with the k there equal to the present $k^M/V(R^M)$.

When $\sigma_{ij} = 0$ for $i \neq j$, Equation 30 reduces to

$$V_i X_i = k^M e_i \qquad i = 1, \ldots, n \tag{31}$$

$$X_i = k^M e_i / V_i \qquad i = 1, \ldots, n. \tag{32}$$

Assuming that the covariance matrix is positive definite, then, $V_i > 0_i$, X_i is positive provided only that excess returns are positive.

In the present uncorrelated case, we can produce an example of two securities, $i =$ 1,2, with $V_1 = V_2$, and $e_1 > e_2$, or $e_1 = e_2$ and $V_1 < V_2$. In the first case, two securities have the same risk but have different excess return. In the second case, they have the same excess return but different risk. So how can it be that the investor is being paid to bear risk? Yet, both of these cases are consistent with market equilibrium. The only difference is that, in accord with Equation 32, the market portfolio will have a greater proportion invested in X_1 than X_2 than it would if $e_1 > e_2$ accompanied $V_1 > V_2$.

Return now to the linear relationship between $Cov(R_i, R^M)$ and e_i in Equation 30, which is equivalent to a linear relationship between β_i and e_i. This holds—not because the investor is paid to bear risk—but because the security with the greater e_i/V_i is a greater fraction of the market. Therefore, the covariance of R_i with R^M is greater because there is more of it in the market.

In the uncorrelated case, any excess return vector $e = (e_1, \ldots, e_n)$ produces $X_i > 0$ for all i so long as $e_i > 0$ for all i. Markowitz (2008) shows that, for an arbitrary covariance matrix, C, there is a cone of e-vectors that are consistent with $X > 0$. This cone has a nonempty interior so long as C is nonsingular. In particular, two securities $i = 1, 2$ can have the same covariance structure $V_1 = V_2$ and $\sigma_{1j} = \sigma_{2j}$ for i and $j \neq 1$ or 2 but not have $e_1 = e_2$. Therefore, it cannot be that the CAPM investor is paid to bear risk because the σ_{ij} structure is the complete risk description in the CAPM.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Allais M. 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'ecole Americaine. *Econometrica* 21:503–46

Barber B, Odean T. 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. *J. Finance* 55(2):773–806

Brinson GP, Hood RL, Beebower GL. 1986. Determinants of portfolio performance. *Financ. Anal. J.* July-Aug:39–44

Brinson GP, Singer BD, Beebower GL. 1991. Determinants of portfolio performance II: an update. *Financ. Anal. J.* 47:40–48

Dantzig GB. 1963. *Linear Programming and Extensions*. Princeton, NJ: Princeton Univ. Press

Dexter AS, Yu JNW, Ziemba WT. 1980. Portfolio selection in a lognormal market when the investor has a power utility function: computational results. In *Stochastic Programming*, ed. MAH Dempster, pp. 507–23. New York: Academic

Ederington LH. 1995. Mean-variance as an approximation to expected utility maximization: semi ex-ante results. In *Advances in Financial Economics*, Vol. 1, ed. M Hirschey, W Marr. Stamford, CT: JAI

Fama EF, French KR. 1992. The cross-section of expected stock returns. *J. Finance* 67(2):427–65

Friedman M, Savage L. 1948. The utility analysis of choices involving risk. *J. Polit. Econ.* 56(4):279–304

Grauer RR. 1986. Normality, solvency, and portfolio choice. *J. Financ. Quant. Anal.* 21(3):265–78

Hlawitschka W. 1994. The empirical nature of Taylor-series approximations to expected utility. *Am. Econ. Rev.* 84(3):713–19

Ibbotson RG, Kaplan PD. 2000. Does asset allocation policy explain 40%, 90% or 100% of performance? *Financ. Anal. J.* 56:26–33

Jacobs BI, Levy KN. 2000. *Equity Management: Quantitative Analysis for Stock Selection*. New York: McGraw-Hill

Jacobs BI, Levy KN, Markowitz H. 2005. Portfolio optimization with factors, scenarios, and realistic short positions. *Oper. Res.* 53(4):586–99

Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47(2):263–91

Kontek K. 2009. *Absolute vs. relative notion of wealth changes*. Work. Pap., Astral Invest. **http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1474229**

Kroll Y, Levy H, Markowitz HM. 1984. Mean variance versus direct utility maximization. *J. Finance* 39(1):47–56

Levy H, Levy M. 2004. Prospect theory and mean-variance analysis. *Rev. Financ. Stud.* 17(4):1015–41

Levy H, Markowitz HM. 1979. Approximating expected utility by a function of mean and variance. *Am. Econ. Rev.* 69(3):308–17

Lintner J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev. Econ. Stat.* 51:222–24

Machina MJ. 1987. Choice under uncertainty: problems solved and unsolved. *Econ. Perspect.* 1(1):121–54

Markowitz HM. 1952a. Portfolio selection. *J. Finance* 7(1):77–91

Markowitz HM. 1952b. The utility of wealth. *J. Polit. Econ.* 60:152–58

Markowitz HM. 1959. *Portfolio Selection: Efficient Diversification of Investments*. Cambridge, MA: Basil Blackwell. 2nd ed.

Markowitz HM. 2005. Market efficiency: a theoretical distinction and so what? *Financ. Anal. J.* 61(5):17–30

Markowitz HM. 2008. CAPM investors do not get paid for bearing risk. *J. Port. Manag.* 34:91–94

Markowitz HM, Todd P. 2000. *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. New Hope, PA: Fabozzi & Assoc.

Markowitz HM, Usmen N. 1996a. The likelihood of various stock market return distributions, part 1: principles of inference. *J. Risk Uncertain.* 13:207–19

Markowitz HM, Usmen N. 1996b. The likelihood of various stock market return distributions, part 2: empirical results. *J. Risk Uncertain.* 13:221–47

Marschak J. 1950. Rational behavior, uncertain prospects, and measurable utility. *Econometrica* 18:111–41

Menger K. 1967. [1934]. The role of uncertainty in economics. In *Essays in Mathematical Economics in Honor of Oskar Morgenstern*, ed. M Shubik. Princeton, NJ: Princeton Univ. Press

Merton RC. 1990. *Continuous Time Finance*. New York: Blackwell

Morningstar, Inc. 2010. *Ibbotson Stocks, Bonds, Bills and Inflation. Classic Yearbook*. Chicago, IL: Morningstar

Pflug GC. 2000. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization: Methodology and Applications*, ed. S Uryasev. Dordrecht: Kluwer Acad.

Pulley LM. 1981. A general mean-variance approximation to expected utility for short holding periods. *J. Financ. Quant. Anal.* 16:361–73

Pulley LM. 1983. Mean-variance approximations to expected logarithmic utility. *Oper. Res.* 31(4):685–96

Rockafellar RT, Uryasev S. 2000. Optimization of conditional value-at-risk. *J. Risk* 2(3):21–41

Rosenberg B. 1974. Extra-market components of covariance in security returns. *J. Financ. Quant. Anal.* 9(2):263–73

Roy AD. 1952. Safety first and the holding of assets. *Econometrica* 20:431–49

Savage LJ. 1954. *The Foundations of Statistics*. New York: Dover. 2nd rev. ed.

Simaan Y. 1987. *Portfolio selection and capital asset pricing for a class of non-spherical distributions of asset returns*. PhD thesis. Baruch College, New York

Sharpe WF. 1963. A simplified model for portfolio analysis. *Manag. Sci.* 9(2):277–93

Sharpe WF. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *J. Finance* 19(3):425–42

Statman M. 2000. The 93.6% question of financial advisors. *J. Invest.* 9:16–20

Thaler R, Johnson E. 1990. Gambling with the house money and trying to break even: the effects of prior outcomes on risky choice. *Manag. Sci.* 36(6):643–60

Vander Weide JH. 2010. Principles for lifetime portfolio selection: lessons from portfolio theory. In *Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques*, ed. JB Guerard Jr. New York: Springer

Von Neumann J, Morgenstern O. 1944. *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton Univ. Press. 3rd ed.

Young WE, Trent RH. 1969. Geometric mean approximation of individual security and portfolio performance. *J. Financ. Quant. Anal.* 4:179–99

Annual Review of
Financial Economics

Volume 2, 2010

Contents

Errata

An online log of corrections to *Annual Review of Financial Economics* articles may be found at http://financial.annualreviews.org