FINANCIAL ANALYSTS JOURNAL®
Reflections

Market Efficiency: A Theoretical Distinction and So What?

Harry M. Markowitz

The capital asset pricing model (CAPM) is an elegant theory. With the aid of some simplifying assumptions, it comes to dramatic conclusions about practical matters, such as how to choose an investment portfolio, how to forecast the expected return of a security or asset class, how to price a new security, or how to price risky assets in a merger or acquisition.

The CAPM starts with some assumptions about investors and markets and deduces its dramatic conclusions from these assumptions. First, it assumes that investors seek mean–variance efficient portfolios; in other words, it assumes that investors seek low volatility and high return on average. Different investors may have different trade-offs between these two, depending on their aversion to risk. Second, the CAPM assumes that taxes, transaction costs, and other illiquidities can be ignored for the purposes of this analysis. In effect, it assumes that such illiquidities may impede the market's approach to the CAPM solution but do not change the general tendency of the market. A third CAPM assumption is that all investors have the same predictions for the expected returns, volatilities, and correlations of securities. This assumption is usually not critical.[1] Finally, the CAPM makes assumptions about what portfolios the investor can select. The original Sharpe (1964)–Lintner (1965) CAPM considered long positions only and assumed that the investor could borrow without limit at the risk-free rate. From this assumption, and the three preceding assumptions, one can deduce conclusions of the sort outlined in the first paragraph.

The assumption that the investor can borrow without limit is crucial to the Sharpe–Lintner model's conclusions. As illustrated later in this article, if we accept the other three CAPM assumptions but assume limited (or no) borrowing, the Sharpe–Lintner conclusions no longer follow. For example, if the four premises of the Sharpe–Lintner original CAPM were true, then the "market portfolio"—a portfolio whose amounts invested are proportional to each security's market capitalization—would be an efficient portfolio. We could not find a portfolio with greater return (on average) without greater volatility. In fact, if the four premises of the Sharpe–Lintner original CAPM were true, the market portfolio, plus perhaps borrowing and lending, would be the *only* efficient portfolio. If, however, we assume the first three premises of the Sharpe–Lintner CAPM but take into account the fact that investors have limited borrowing capacity, then it no longer follows that the market portfolio is efficient. As this article will illustrate, this inefficiency of the market portfolio could be substantial and it would not be arbitraged away even if some investors could borrow without limit.

When one clearly unrealistic assumption of the capital asset pricing model is replaced by a real-world version, some of the dramatic CAPM conclusions no longer follow.

Harry M. Markowitz is president of Harry Markowitz Company, San Diego, California.

Before the CAPM, conventional wisdom was that some investments were suitable for widows and orphans whereas others were suitable only for those prepared to take on "a businessman's risk." The CAPM convinced many that this conventional wisdom was wrong; the market portfolio is the proper mix among risky securities for everyone. The portfolios of the widow and businessman should differ only in the amount of cash or leverage used. As we will see, however, an analysis that takes into account limited borrowing capacity implies that the pre-CAPM conventional wisdom is probably correct.

An alternate version of the CAPM speaks of investors holding short as well as long positions. But the portfolios this alternate CAPM permits are as unrealistic as those of the Sharpe–Lintner CAPM with unlimited borrowing. The alternate CAPM assumes that the proceeds of a short sale can be used, without limit, to buy securities long. For example, the alternate CAPM assumes that an investor could deposit $1,000 with a broker, short $1,000,000 worth of Stock A, then use the proceeds and the original deposit to buy $1,001,000 of Stock B. The world does not work this way.

Like the original CAPM, the alternate CAPM implies that the market portfolio is an efficient portfolio, although not the only one (as in the original CAPM). If one takes into account real-world constraints on the holding of short and long positions, however, the efficiency of the market portfolio no longer follows, as will be illustrated.

Both the original CAPM, with unlimited borrowing, and the alternate CAPM, with unrealistic short rules, imply that the expected return of a stock depends in a simple (linear) way on its beta, and only on its beta. This conclusion has been used for estimating expected returns, but it has lost favor for this use because of poor predictive results. It is still used routinely in "risk adjustment," however, for valuing assets and analyzing investment strategies on a "risk-adjusted basis." I will show here that the conclusion that expected returns are linear functions of beta does not hold when real-world limits on permitted portfolio holdings are introduced into the CAPM. This discussion will call into question the frequent use of beta in risk adjustment.

I will discuss the assumptions and conclusions of the CAPM formally and then illustrate the effect on CAPM conclusions of varying the CAPM assumptions concerning the investor's constraint set. Afterward, I will sketch how the points illustrated in the simple examples generalize to more complex cases. Finally, I will discuss the implications of the analysis for financial theory, practice, and pedagogy.

A Distinction

We should distinguish between the statement that "the market is efficient," in the sense that market participants have accurate information and use it correctly to their benefit, and the statement that "the market portfolio is an efficient portfolio." Under some conditions, the former implies the latter. In particular, if one makes the following assumptions,

A1. transaction costs and other illiquidities can be ignored (as I will do throughout this article),

A2. all investors hold mean–variance efficient portfolios,

A3. all investors hold the same (correct) beliefs about means, variances, and covariances of securities, and—*in addition*—

A4. every investor can lend all she or he has or can borrow all she or he wants at the risk-free rate,

then Conclusion 1 follows:

C1. The market portfolio is a mean–variance efficient portfolio.

C1 also follows if A4 is replaced by A4′:

A4′. investors can sell short without limit and use the proceeds of the sale to buy long positions.

In particular, A4′ says that any investor can deposit $1,000 with a broker, short $1,000,000 worth of one security, and buy long $1,001,000 worth of another security.

Neither A4 nor A4′ is realistic. Regarding A4, when an investor borrows, not only does the investor pay more than when the U.S. government borrows, but (a point of equal or greater importance here) the amount of credit extended is limited to what the lender believes the borrower has a reasonable probability of repaying. Regarding A4′, if the investor deposits $1,000 with a broker, Federal Reserve Regulation T permits the investor to buy a $2,000 long position *or* take on a $2,000 short position *or* take on a $1,000 long and a $1,000 short position, but it does not allow an unlimited amount short plus the same unlimited amount long, as assumed in A4′.

If one replaces A4 or A4′ with a more realistic description of the investor's investment constraints, then C1 usually no longer follows; even though all investors share the same beliefs and each holds a mean–variance efficient portfolio, the market portfolio need not be an efficient portfolio. This departure from efficiency can be quite substantial. In fact, the market portfolio can have almost *maximum* variance among feasible portfolios with the same expected value rather than *minimum* such variance; that is, the market portfolio can be about as *in*efficient as a feasible portfolio can get (see Chapter 11 of Markowitz 1987 or Markowitz and Todd 2000).

In addition to C1, A1 through A4 (or A1 through A4′) imply

C2. In equilibrium, the expected return for each security depends only on its beta (the regression of its returns against the return on the market). This relationship between the security's expected return and its beta is a simple, linear relationship.

C2 is the basis for the CAPM's prescriptions for risk adjustment and asset valuation. Like the first conclusion, C2 does not follow from assumptions A1 through A3 if A4 (or A4′) is replaced by a more realistic description of the investor's investment constraints.

Often, financial research attempts to determine "market efficiency" by testing whether C2 holds. But the failure of C2 to hold empirically does not prove that the market is not efficient in the general sense of possessing correct information and using it advantageously. Nor does the failure of C2 to hold empirically prove that the market is not efficient in the narrower sense of A2 and A3—namely, that participants hold mean–variance efficient portfolios in terms of commonly held correct beliefs. I will not argue here that A2 and A3 are true—or argue that they are false. I argue only that, in the absence of either A4 or A4′, the empirical refutation of C2 is not an empirical refutation of A1 through A3.[2]

Example. In this first example, I assume that investors cannot sell short or borrow (but I note subsequently that the same results hold if investors *can* borrow limited amounts or *can* sell short but are subject to Reg T or some similar constraint). The example assumes A1 through A3; that is, it ignores taxes, transaction costs, and other illiquidities; it assumes that all investors have the same beliefs about the means, variances, and covariances of security returns; and it assumes that each investor holds a portfolio that is mean–variance efficient in terms of these beliefs.

This example consists of long positions in three risky securities with the expected returns and standard deviations shown in **Table 1**. To keep things simple, we will assume that returns are uncorrelated. However, the results also hold for correlated returns.

Table 1. Expected Returns and Standard Deviations of Three Risky Securities

Security	Expected Return	Standard Deviation
1	0.15%	0.18%
2	0.10	0.12
3	0.20	0.30

Let X_1, X_2, and X_3 represent the fraction of her wealth that some investor invests in, respectively, Securities 1, 2, and 3. Assume that the investor can choose any portfolio that meets the following constraints:

$$X_1 + X_2 + X_3 = 1.0 \tag{1}$$

and

$$X_1 \geq 0, X_2 \geq 0, X_3 \geq 0. \tag{2}$$

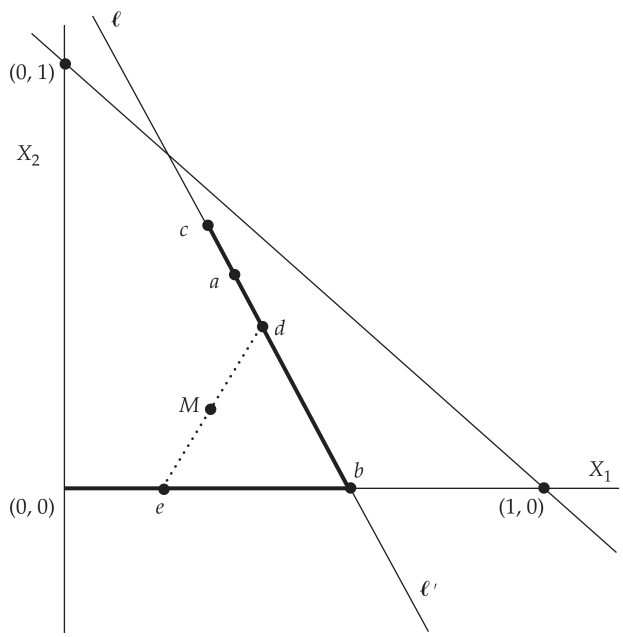
The first of these is a budget equation; the second is a requirement that none of the investments be negative. We will contrast the efficient set and market portfolio we get with Budget Equation 1 and Nonnegativity Requirement 2 as constraints with the set and portfolio we get if we assume A4′ (that is, if we assume that Budget Equation 1 is the only constraint). In **Figure 1**, X_1—the fraction invested in Security 1—is plotted on the horizontal axis; X_2—the fraction in Security 2—is plotted on the vertical axis; and X_3—the fraction invested in the third security—is given implicitly by the relationship

$$X_3 = 1 - X_1 - X_2. \tag{3}$$

Figure 1 should be thought of as extended without limits in all directions. Every point (portfolio) on this extended page is feasible according to assumption A4′. For example, the point with $X_1 = 93$ and $X_2 = -106$ (therefore, $X_3 = 14$ according to Equation 3) is feasible according to assumption A4′ because it satisfies Equation 1. It is not feasible when Budget Equation 1 *and* Nonnegativity Requirement 2 are required because it does not satisfy $X_2 \geq 0$.

The only points (portfolios) in Figure 1 that satisfy Budget Equation 1 and Nonnegativity Requirement 2 are on and in the triangle whose vertices are the points (1,0), (0,1), and (0,0). The first

Figure 1. Efficient Sets with and without Nonnegativity Constraints



of these points represents an undiversified portfolio with 100 percent invested in Security 1 (X_1 = 1.0); the second, a 100 percent investment in Security 2 (X_2 = 1.0); the third, a 100 percent investment in Security 3 (X_3 = 1.0). The diagonal side of the triangle connecting points (1,0) and (0,1) includes investments in Securities 1 and 2 but not Security 3; the horizontal side connecting (0,0) and (1,0) has investments in Securities 1 and 3 but not in Security 2; the side connecting (0,0) and (0,1) has X_1 = 0. Points within the triangle represent portfolios with positive investments in all three securities. The vertices, sides, and interior of the triangle all meet Budget Equation 1 and Nonnegativity Requirement 2.

If assumption A4′ holds (therefore, Budget Equation 1 is the only constraint), then all the portfolios with the least standard deviation for various levels of expected return lie on the straight line labeled $\ell\ell'$ in Figure 1. Because two points determine a line, we know the whole line if we know two points on it. One point on the line is the portfolio that minimizes standard deviation among all portfolios on the extended page (i.e., among all portfolios that satisfy Equation 1). When returns are uncorrelated, this risk-minimizing portfolio satisfies

$$X_1 = \frac{K_c}{V_1}, \tag{4a}$$

$$X_2 = \frac{K_c}{V_2}, \tag{4b}$$

and

$$X_3 = \frac{K_c}{V_3}, \tag{4c}$$

where V_1, V_2, V_3 are the variances (standard deviations squared) of the three securities and K_c is chosen so that Equation 1 is satisfied; that is,

$$K_c = \frac{1}{\left(1/V_1\right) + \left(1/V_2\right) + \left(1/V_3\right)}. \tag{5}$$

Thus, when returns are uncorrelated, the variance-minimizing portfolio is always within the triangle. For the current example,

$X_1 = 0.28,$

$X_2 = 0.62,$

and

$X_3 = 0.10.$

This point is the point labeled "*c*" in Figure 1.

When returns are uncorrelated, another point on the line that minimizes portfolio variance for various levels of portfolio expected return is

$$X_1 = \frac{K_a E_1}{V_1}, \tag{6a}$$

$$X_2 = \frac{K_a E_2}{V_2}, \tag{6b}$$

and

$$X_3 = \frac{K_a E_3}{V_3}, \tag{6c}$$

where E_1, E_2, and E_3 are the expected returns of the three securities and K_a is chosen to satisfy Equation 1. In our example, this is the portfolio

$X_1 = 0.34,$

$X_2 = 0.50,$

and

$X_3 = 0.16.$

It is the point labeled "*a*" in Figure 1.

If we continue to assume Budget Equation 1 as the only constraint, all points on the straight line through *a* and *c* minimize portfolio variance for various levels of portfolio expected return. However, not all these points are *efficient* portfolios. Efficient portfolios are those encountered if we start at *c* and move continuously in the direction of *a*, and beyond, without stop. As we move away from c in this direction, portfolio expected return, E_P, and portfolio variance, V_P, increase. All portfolios encountered provide minimum V_P for the given E_P—*or greater* E_P—among all portfolios that satisfy Budget Equation 1. In contrast, if we start at c and move in the other direction, we do not encounter efficient portfolios (other than c) because V_P increases but E_P decreases. The same V_P but greater E_P can be found elsewhere on $\ell\ell'$.

Thus, in this example, if Budget Equation 1 is the only constraint, the set of efficient portfolios is the "ray" that starts at c and moves in a straight line through a and beyond.

As one moves on the line $\ell\ell'$ in the direction of a and beyond, at some point the line $\ell\ell'$ leaves the triangle. In the present example, this is the point labeled "b" in Figure 1, with

$$X_1 = 0.58,$$

$$X_2 = 0.00,$$

and

$$X_3 = 0.42.$$

Portfolio b still satisfies the constraints (Budget Equation 1 and Nonnegativity Requirement 2), but points beyond b on the line $\ell\ell'$ no longer satisfy Nonnegativity Requirement 2 because they violate the requirement that $X_2 \geq 0$. Beyond point b, therefore, the efficient set when Budget Equation 1 *and* Nonnegativity Requirement 2 are required departs from the efficient set when Budget Equation 1 only is required.

At point b, investment in Security 2 is zero ($X_2 = 0$). For efficient portfolios with higher expected return, the efficient set moves along the horizontal edge of the triangle, from b to (0,0), where an undiversified portfolio is invested only in Security 3 ($X_3 = 1$), the security with the highest expected return in the example.

We will see that, quite generally, a set of mean–variance efficient portfolios is "piecewise linear"; that is, it is made up of one or more straight-line segments that meet at points called "corner portfolios." When Equation 1 is the only constraint, the efficient set contains only one corner portfolio—namely, point c in Figure 1—and only one line "segment"—namely, the segment that starts at c and moves without end in the direction of increasing E_p. When nonnegativity constraints are imposed, the set of efficient portfolios typically has more than one segment and more than one corner portfolio. In Figure 1, this set of efficient portfolios consists of two line segments connecting three corner portfolios—c, b, and (0,0).

The Two-Fund Separation Theorem. The fact that two points determine a line is known in financial theory as the "two-fund separation theorem." In particular, all the portfolios on $\ell\ell'$ in Figure 1 can be obtained by (positive or negative) investments in portfolios a and c subject only to the constraint

$$X_a + X_c = 1, \qquad (7)$$

where X_a and X_c are the "fractions" of the portfolio allocated to, respectively, subportfolios a and c. Note that Equation 7 permits the investor to short one portfolio and use the proceeds to invest more than 100 percent in the other portfolio. If both X_a and X_c are positive, then the resulting portfolio lies within the interval connecting a and c in Figure 1. If X_c is negative, then $X_a > 1$ and the resulting portfolio lies outside the interval, beyond a. Similarly, if $X_a < 0$ and $X_c > 1$, the portfolio lies outside the interval beyond c.

What is true in particular on $\ell\ell'$ is true in general for any two distinct points on any line in portfolio space. All points on the line can be represented by investments X_a and X_c in two distinct subportfolios on the line, where X_a and X_c satisfy Equation 7. I will use this relationship between points and lines several times.

The Market Portfolio. Consider a market in which investors must satisfy Budget Equation 1 and Nonnegativity Requirement 2. I show in the next section and in Appendix A that—in this case—beliefs about means, variances, and covariances that imply the efficient set in Figure 1 are consistent with market equilibrium.

Assume there are two types of investors in this market: cautious investors who select the portfolio at d = (0.40, 0.37) in Figure 1 and aggressive investors who select the portfolio at e = (0.20, 0.00). Similar conclusions would be reached if we specified two other portfolios as long as one of the portfolios were on one of the segments and the other portfolio were on the other segment of the efficient set. Similar conclusions would also be reached if there were more than two types of investors as long as some were on one segment and some on the other.

According to the two-fund separation theorem, the market portfolio lies on the straight line connecting *d* and *e* [for example, at $M = (0.30, 0.19)$]. The market is efficient, in that each participant holds an efficient portfolio, but note that the *market portfolio*, *M*, is not an efficient portfolio. It is not on either segment of the efficient set when Budget Equation 1 and Nonnegativity Requirement 2 are the constraints (nor is it, incidentally, on the ray that is the efficient set when Budget Equation 1 only is the constraint).

A Simple Market. The preceding shows that if investors selected portfolios subject to the constraints of Budget Equation 1 and Nonnegativity Requirement 2, all held the beliefs in Table 1, and some preferred portfolios on one segment of the efficient set and others preferred a portfolio on the other, then the market portfolio would not be a mean–variance efficient portfolio. This section shows that means, variances, and covariances that imply Figure 1 are consistent with economic equilibrium when shorting and borrowing are unavailable.

Imagine an economy in which the inhabitants live on coconuts and the produce of their own gardens. The economy has three enterprises, namely, three coconut farms. Once a year, a market convenes to trade the shares of the three coconut farms. Each year, the resulting prices of shares turn out to be the same as those of preceding years because the number of people with given endowments and risk aversion is the same each year (perhaps because of overlapping generations rather than immortal participants). Thus, the only source of uncertainty of return is the dividend each stock pays during the year—which is the stock's pro rata share of the farm's production.

It is shown in Appendix A that means, variances, and covariances of coconut production exist that imply the efficient set in Figure 1—or any other three-security efficient set that we cite. If we insist that coconut production be nonnegative, it may be necessary to add a constant to all expected returns (the same constant to each). Doing so will increase the expected returns of each portfolio but not change the set of efficient portfolios. It is then possible to find a probability distribution of coconut production, with production always nonnegative, for the given (slightly modified) means, variances, and covariances and, therefore, for the given set of efficient portfolios.

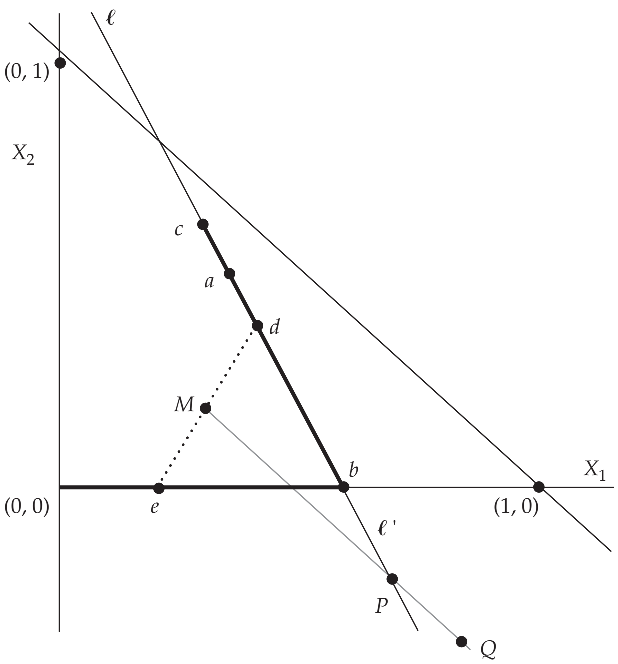
With such a probability distribution of returns, the market is rational, in the sense that each participant knows the true probability distribution of returns and each seeks and achieves mean–variance efficiency. Nevertheless, in contrast to the usual CAPM conclusion, the market portfolio is not an efficient portfolio. It follows that there is no representative investor, because no investor wants to hold the market portfolio. Also, as we will see in a subsequent section, expected returns are not linearly related to betas.

Arbitrage. Suppose that most investors are subject to Nonnegativity Requirement 2 but that one investor can short, in the CAPM sense—that is, is subject only to Budget Equation 1. (Perhaps the CAPM investor has surreptitious access to a vault containing stock certificates that he or she can "borrow" temporarily without posting collateral.) Would this CAPM investor arbitrage away the inefficiency in the market portfolio?

If there were a Portfolio *P* on BB' that beat Market Portfolio *M* with certainty, then the CAPM investor could short any amount of *M*, use the proceeds to buy *P*, and make an arbitrarily large gain with certainty. But *P* does not beat *M* with certainty; it simply offers a better probability distribution. In fact, the investor with Equation 1 as the only constraint is better off picking a point on BB' and ignoring *M*. **Figure 2** illustrates this idea. If *P* is any point *on* the line BB' and *M* is any point *off* the line BB', then according to the two-fund separation theorem, the portfolio produced by shorting *M* and using the proceeds (plus the original "$1") to buy *P* lies on the straight line connecting *M* and *P*. Specifically, it lies on the far side from *M*, beyond *P*, such as *Q* in Figure 2. But Portfolio *Q* is not efficient for the investor with Equation 1 as the only constraint. Some Portfolio *R* on BB' (not shown in Figure 2) supplies a higher mean and lower variance.

Now that we have seen that an investor subject only to Equation 1 will choose a portfolio from BB' without regard to the market portfolio, let us consider market equilibrium when some investors are subject to Equation 1 only and some to Equation 1 and Nonnegativity Requirement 2. Suppose that, as in Figure 1, the average holdings (weighted by investor wealth) of investors subject to Budget Equation 1 and Nonnegativity Requirement 2 is the point *M*. It would be the market portfolio if these were the only investors. Suppose further that the wealth-weighted average of the one or more investors subject only to Equation 1 is

Figure 2. Effect of Trying to Arbitrage an Inefficient Market Portfolio



point P in **Figure 3**. As in Figure 2, P must lie on BB', whereas M typically lies off BB'. When both types of investors are present, the market portfolio lies on the straight line between the two averages, M and P, such as point M^a in Figure 3. The position of M^a depends on the relative wealth of the two types of investors, but in any case, it is off BB'; therefore, it is not efficient for investors subject to Equation 1 only.

Whether it is efficient for investors subject to both Budget Equation 1 and Nonnegativity Requirement 2 is a more complicated story. The portfolios M^a, M^b, and M^c lie on the straight line connecting P and M. Portfolio M^c cannot be a market equilibrium because it implies a negative total demand for Security 2. If M^b is the market portfolio, then the market portfolio is efficient for investors with Budget Equation 1 and Nonnegativity Requirement 2 as constraints but there is zero net demand for shares of Security 2. For an equilibrium with positive net demand for all three securities, the market must be within the constraint triangle, as M^a is in Figure 3. But such a combination of M and P is inefficient for investors subject to Budget Equation 1 and Nonnegativity Requirement 2, as well as for those subject only to Budget Equation 1.

Expected Returns and Betas. If Assumptions 1–4 or 1–4′ are true, then Conclusion 2 follows: Expected returns are linearly related to the betas of each security. That is, for some choice of numbers a and b, the following three equations hold:

$$E_1 = a + b\beta_1, \tag{8a}$$

$$E_2 = a + b\beta_2, \tag{8b}$$

and

$$E_3 = a + b\beta_3, \tag{8c}$$

where β_i is the coefficient of regression of the return on the ith security against the return on the market. But these equations do not necessarily hold if A1–A3 are true but neither A4 nor A4′ is true.

In particular, C2 will typically not be true if investors satisfy A1–A3 but are subject to Equation 1 and Nonnegativity Requirement 2 as constraints. I will illustrate this statement in terms of the three-security example in Figure 3.

The first column of **Table 2** shows the fraction P_i of security i in Portfolio P on the BB' line in Figure 3. The second column states the covariance between each security and P. Given our current assumption that the returns on the three securities are uncorrelated, the covariance between security i and Portfolio P depends only on how much of the security is in the portfolio, and it is given by the formula

$$\text{cov}(R_i, P) = P_i V_i, \text{ for } i = 1, 2, 3. \tag{9}$$

Figure 3. Market Portfolios with and without Nonnegativity Constraints

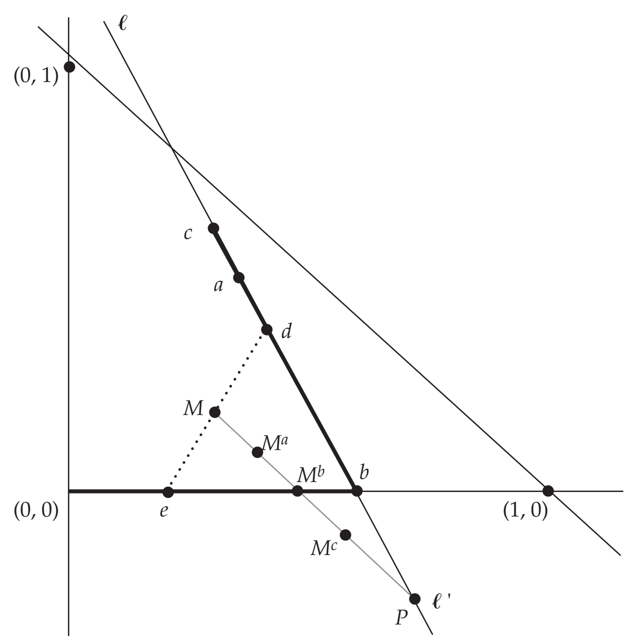


Table 2. Three Risky Securities in Portfolio *P* of Figure 3

Security	Percent in *P*	$\text{cov}_{i,P} = P_i V_i$	$\beta_{i,P}$
1	0.70%	0.0227	0.52
2	–0.25	–0.0036	–0.08
3	0.55	0.0495	1.12

Note: $\text{var}(P) = 0.0440$; $\beta_{i,P} = \text{cov}_{i,P}/\text{var}(P)$.

Table 3. Three Risky Securities in Market Portfolio *M* of Figure 3

Security	Percent in *M*	$\text{cov}_{i,M} = M_i V_i$	$\beta_{i,M}$
1	0.30%	0.0097	0.36
2	0.19	0.0027	0.10
3	0.51	0.0459	1.71

Note: $\text{var}(M) = 0.0268$; $\beta_{i,M} = \text{cov}_{i,M}/\text{var}(M)$.

The beta of any security return regressed against any Portfolio *P* is defined to be

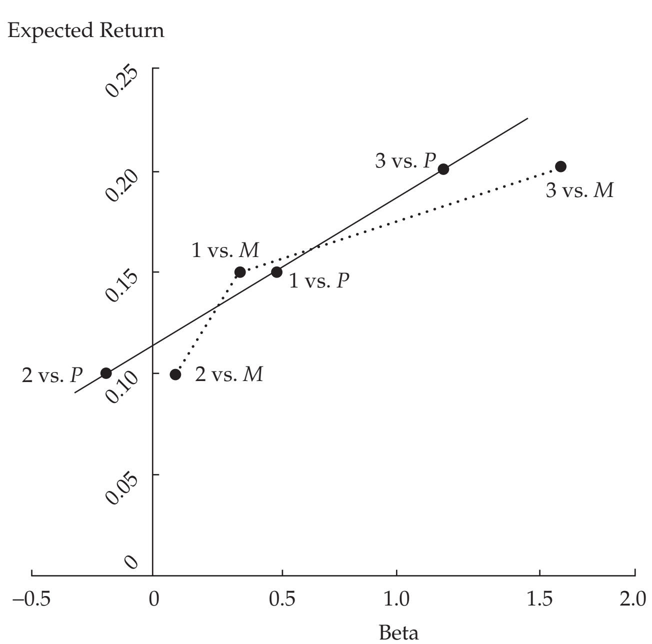
$$\beta_{i,P} = \frac{\text{cov}_{i,P}}{\text{var}(P)}. \tag{10}$$

These betas are listed in the last column of Table 2.

Similarly, **Table 3** shows the fraction held in Market Portfolio *M*, the covariance between each security and Portfolio *M*, and the beta of each security return regressed against the return on *M*, where *M* in Figure 3 is also the market portfolio in Figure 1.

In **Figure 4**, the points labeled 1 vs. *P*, 2 vs. *P*, and 3 vs. *P* show expected return on the vertical axis against $\beta_{i,P}$ plotted on the horizontal axis. The points labeled 1 vs. *M*, 2 vs. *M*, and 3 vs. *M* show the same expected returns plotted against $\beta_{i,M}$. The three observations for each case are connected by lines. We see that the three points that represent expected returns and betas-versus-*P* lie on a single straight line whereas the three points representing expected returns and betas-versus-*M* do not lie on a straight line. The implication is that there is a linear relationship between expected returns and betas-versus-*P* but no such relationship between expected returns and betas-versus-*M*. In other words, for some choice of *a* and *b*, Equation 8 holds if the betas in Equation 8 are from regressions against *P* but no such *a* and *b* choice exists when the betas are from regressions against *M*. More generally, if Market Portfolio *M* is any point on BB′, then a linear relationship exists between expected return and beta. In contrast, if *M* is any point off BB′, there is no such relationship (see Roll 1977; Markowitz 1987; Markowitz and Todd).

Figure 4. Relationship between Expected Returns and Betas versus an Efficient and an Inefficient Market Portfolio



Limited Borrowing. In this section, I introduce a risk-free asset into the discussion. The Sharpe–Lintner CAPM assumes A1–A4 including unlimited borrowing at the risk-free rate. These assumptions imply that the market portfolio is a mean–variance efficient portfolio and that expected returns are linearly related to betas against the market portfolio. In this section, I illustrate that this conclusion no longer follows if borrowing is either not permitted or permitted but limited.

To illustrate this idea, the example in Table 1 is modified so that Security 3 now has 0 variance and a (risk-free) return of r_0 = 3 percent, as shown in **Table 4**. We continue to assume the budget constraint (Equation 1) and

$$X_1 \geq 0 \text{ and } X_2 \geq 0. \tag{11a}$$

$X_3 > 0$ represents lending at the risk-free rate; $X_3 < 0$ represents borrowing at the same rate. Prohibited borrowing would be represented by the constraint

$$X_3 \geq 0. \tag{11b}$$

Borrowing limited to, for example, the equity in the account would be represented by

$$X_3 \geq -1.0. \tag{11c}$$

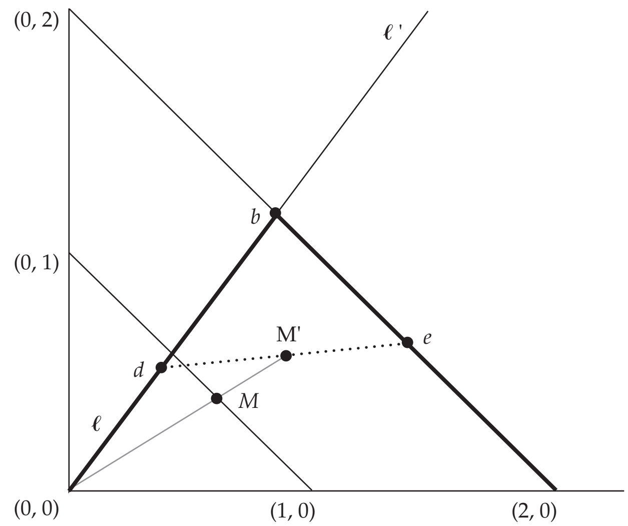
Unlimited borrowing would be represented by the constraints of Budget Equation 1 and Nonnegativity Requirement 11a, with no constraint on X_3.

Table 4. Expected Returns and Standard Deviations of Three Securities Including Cash

Security	Expected Return	Standard Deviation
1	0.15%	0.18%
2	0.10	0.12
3	0.03	0.00

In **Figure 5**, as in Figure 1, the horizontal axis represents X_1, the fraction of the portfolio invested in Security 1; the vertical axis represents X_2, the fraction invested in Security 2. As before, X_3 is given implicitly by Equation 3. If borrowing is forbidden, then the set of feasible portfolios is, as before, on and in the triangle with vertices (0,0), (1,0), and (0,1). If no more than 100 percent borrowing is permitted, the set of feasible portfolios is the points on and in the triangle whose vertices are (0,0), (2,0), and (0,2). If unlimited borrowing is permitted, the set of feasible portfolios is the entire positive quadrant.

Figure 5. Market Portfolio when Borrowing Permitted but Limited



In our example assuming uncorrelated returns, when borrowing is unconstrained, the set of efficient portfolios is the set of portfolios that satisfies

$$X_1 = \frac{h(E_1 - r_0)}{V_1} \tag{12a}$$

and

$$X_2 = \frac{h(E_2 - r_0)}{V_2} \tag{12b}$$

for any zero or positive choice of h. This line is the ray that starts at the origin (0,0)—the all-cash portfolio—and proceeds into the positive quadrant along line BB′ in Figure 5 passing through the point (0.43, 0.57) for the example in Table 4. When borrowing is not limited, the efficient set proceeds along BB′ without bounds. If investors cannot borrow more than 100 percent of equity, then the efficient set cannot go beyond the line connecting (2,0) and (0,2); that is, it cannot go beyond point b = (0.86, 1.14). From that point, under Nonnegativity Requirement 11c, the efficient set moves along the line connecting (0,2) and (2,0) until it reaches the point (2,0), representing the portfolio that is 200 percent invested in the highest-yielding security, namely, Security 1 in the present example.

Suppose some investors choose portfolio d = (0.39, 0.51) on one segment in Figure 5 and all others choose portfolio e = (1.40, 0.60) on the other segment.

Then, "the market"—including cash or borrowing—is a point between them, such as M'. Portfolio M is M' "normalized" so that the "market portfolio" adds up to 100 percent. Neither M nor M' is an efficient portfolio. Nor is there a linear relationship between expected returns and betas regressed against either M or M'. Such a relationship exists only if M is on $\mathbb{B}\mathbb{B}'$. The fact that the market is inefficient implies that there is no representative investor. No rational investor holds either M or M'.

Generalizations

Mean–variance efficient sets are computed in practice for models ranging in size from toy problems with two, three, or four assets to small problems with a dozen or so asset classes to large problems containing thousands of securities. To calculate an efficient frontier, the "critical line algorithm" (CLA) accepts as inputs any vector of expected return estimates, any matrix of covariance estimates (even a singular covariance matrix), and any linear equality or inequality constraints on the choice of portfolio (such as upper bounds on individual security holdings, sums of security holdings, or weighted sums of security holdings). From these inputs, the CLA produces a piecewise linear set of efficient portfolios. This set of portfolios "looks like" the ones in Figures 1–5, except that now the sets are difficult to draw because a portfolio in an analysis with 1,000 securities requires approximately a 1,000-dimensional space to be plotted. (When portfolio choice is subject to a budget constraint, a 999-dimensional space is sufficient.) Although we cannot plot points on a 999-dimensional blackboard, the basic mathematical properties of points, lines, and efficient sets in 999-dimensional space are the same as those in 2-dimensional space. The diagrams in Figures 1–5, which illustrate these properties, can help our intuition as to the nature of the properties in higher-dimensional spaces.

One property of points and lines that is the same in 999-dimensional space as it is in 2-dimensional space is that two points determine a line. In particular, all portfolios that lie on a straight line in an any-dimensional space may be obtained by investing amounts X_a and X_c in Portfolios P_a and P_c on the line. P_a and P_c may be any two fixed, *different* portfolios on the line. As in the three-security case, X_a and X_c are subject to constraining Equation 7 and either may be negative. If X_a is negative, then $X_c > 1.0$ and Point (Portfolio) P obtained by allocating X_a to P_a and X_c to P_c lies outside the interval connecting P_a and P_c, beyond P_c. The other cases—with $X_c < 0$ or with $X_a \geq 0$ and $X_c \geq 0$—are as described in the discussion of the two-fund separation theorem for the three-security case.

Suppose that there are n securities (for $n = 3$ or 30 or 3,000), that not all expected returns are the same, and that the n securities have a nonsingular covariance matrix. If the only constraint on the choice of portfolio is

$$\sum_{i=1}^{n} X_i = 1, \tag{13}$$

then the portfolios that minimize portfolio variance V_P for various values of portfolio expected return E_P lie on a single straight line $\mathbb{B}\mathbb{B}'$ in $(n-1)$-dimensional portfolio space. Expected return increases as one moves in one direction on this line, decreases in the other direction. The set of efficient portfolios in this case is the ray that starts at the V_P-minimizing portfolio and moves on $\mathbb{B}\mathbb{B}'$ in the direction of increasing E_P. Repeated use of the two-fund separation theorem shows that if all investors hold portfolios somewhere on this ray, the market portfolio will also be on this ray and, therefore, will also be efficient. Thus, the efficiency of the market portfolio when $n = 3$ and Equation 1 is the only constraint generalizes to any n with Equation 13 as the only constraint.

Next, consider an investor subject to a no-shorting constraint,

$$X_i \geq 0,\ i = 1, \ldots, n, \tag{14}$$

as well as a budget constraint (Equation 13). For simplicity, assume that one security has the greatest expected return (albeit, perhaps, just slightly more than the second-greatest expected return). When Budget Equation 13 *and* Nonnegativity Requirement 14 are the constraints, and the only constraints, on portfolio choice, the critical line algorithm begins with the portfolio with highest expected return, namely, the portfolio that is 100 percent invested in the security with highest expected return. The CLA traces out the set of efficient portfolios from top to bottom (i.e., from this portfolio with maximum expected return down to the portfolio with minimum variance). The computation proceeds in a series of iterations. Each iteration computes one piece (one linear segment) of the piecewise linear efficient set. Each successive segment has either one more or one less security than the preceding segment. If the analysis includes a risk-free asset (or, equivalently, risk-free lending), the last segment to be computed (the one with the lowest portfolio mean and variance) is the one and only segment that contains the risk-free asset (Tobin 1958).

This characterization of efficient sets remains true if limited borrowing is allowed, as illustrated in Figure 5. It also remains true when short selling is permitted but is subject to a Reg T or similar constraint (see Jacobs, Levy, and Markowitz, forthcoming). In this case, if no other constraints are included (such as upper bounds on holdings), then short sales subject to Reg T can be modeled by an analysis with $2n + 3$ variables. The first n variables represent long positions; the second n variables represent short positions; and the final three variables represent, respectively, lending, borrowing, and slack in the Reg T constraint. These variables are subject to the following constraints:

$$\sum_{i=1}^{n} X_i + X_{2n+1} - X_{2n+2} = 1, \tag{15a}$$

$$\sum_{i=1}^{2n} X_i + X_{2n+3} = 2, \tag{15b}$$

and

$$X_i \geq 0, \text{ with } i = 1, \ldots, 2n + 3. \tag{15c}$$

The portfolio with maximum expected return typically contains two variables at positive levels (perhaps a short or long position plus borrowing). As in the case without short positions, the CLA traces out the efficient frontier in a series of iterations—each iteration producing one piece of the piecewise linear efficient set, each piece having one more or (occasionally) one less nonzero variable than did the preceding piece.

A great variety of mean–variance efficient sets are computed in practice. For example, some are computed for asset classes; some of these results are then implemented by index funds. Other efficient set analyses are performed at the individual-security level. Among the latter, analyses differ as to which securities constitute "the universe" of securities from which the portfolio optimizer is to select for its portfolios. Some permit short positions; some do not.

For comparability with the classic CAPM, let us assume here that all investors perform their mean–variance analyses in terms of individual securities rather than asset classes, all use the same universe of "all marketable securities," and either all include short sales (subject to a Reg T–like constraint) or all exclude short sales.

Even so, there properly should be a variety of portfolio analyses generating a variety of frontiers. Because different institutions have different liability structures, they properly have different efficient sets of marketable securities. For example, an insurance company or pension fund, with liabilities determined outside the portfolio analysis, should choose portfolios that are efficient in terms of the mean and variance of assets minus liabilities. When different investors properly have different efficient sets, the question of whether the market portfolio is a mean–variance efficient portfolio raises the question: efficient for whom?

For comparability with the CAPM, let us assume that all investors may properly ignore their particular liability structure in computing the efficient frontier; each uses the same mean, variance, and covariance estimates for the same universe of marketable securities; and each is subject to the same constraints. In other words, we assume that they all generate and select portfolios from the same mean–variance efficient frontier.

In tracing out this frontier, the CLA starts at the high end with an undiversified portfolio. It proceeds in a sequence of iterations that generate "lower" segments of the piecewise linear efficient frontier (i.e., segments with lower portfolio mean and lower portfolio variance). Each successive segment adds or deletes one security (or possibly a short position) on the list of active securities. Thus, if the universe consists of, say, 10,000 securities, then if all securities are to be demanded by someone, this universal efficient frontier must contain at least 10,000 segments. If investors have sufficiently diverse risk tolerances, they will choose portfolios on different segments. Some will prefer portfolios on one or another of the typically less diversified high-risk/high-return segments. Others will select portfolios on one or another of the typically more diversified lower-risk segments. The market is an average, weighted by investor wealth, of portfolios selected from these diverse segments. Although it is mathematically possible for this average to accidentally fall on the efficient frontier, such an outcome is extremely unlikely.

Thus, in this world that is like the CAPM but has realistic constraints, the market portfolio is typically not an efficient portfolio. Therefore, there is no representative investor and expected return is not a linear function of regressions of security returns against the market.

So What?

This section presents some implications of the preceding analysis.

So What #1. A frequent explanation of why observed expected returns do not appear to be linearly related to betas is that the measures of market return used in the tests do not measure the true, universal market portfolio that appears in the CAPM. The conclusion is that to test the CAPM, we need to measure returns on a cap-weighted world portfolio. The preceding discussion implies, however, that before spending vast resources on ever finer approximations to returns on this cap-weighted universal portfolio, we should note that CAPM Conclusion 2 (that expected returns are linearly related to betas) is not likely to be true if real-world constraints are substituted for Assumption 4 or Assumption 4′.

So What #2. Traditionally, some investments were thought of as businessmen's risks while others were thought appropriate for widows and orphans. The CAPM, assuming A1–A4, concludes that one and only one portfolio is efficient for all investors. The only difference should be the amount of cash or borrowing with which the portfolio is combined. In contrast, when borrowing is limited and short sales are prohibited or subject to real-world constraints, the composition of the portfolio of risky securities changes radically from one end to the other of the efficient frontier. At the high end, it contains few securities, usually with a predominance of those with high expected return. At the low end, it tends to be more diversified, with a more-than-proportional presence of the less volatile securities. In other words, the high end of the frontier will indeed tend to be dominated by businessman-risk securities, whereas the low end, although perhaps spiced up and diversified with some more-volatile securities, will typically have more than its proportionate share of widow-and-orphan securities.

So What #3. The linear relationship between expected returns and betas (against the market portfolio return) that is implied by the CAPM is the basis for a standard "risk-adjustment" calculation. This calculation is used, for example, to determine which of two projects that a company might pursue would best enhance its stock market value or which of two securities, groups of securities, or investment strategies has performed best. Because the existence of a linear relationship between expected returns and betas is questionable, the reliability of its use in risk adjustment must be questioned.

It might seem at first that the use of the CAPM risk-adjustment formula is indispensable for decisions like those I just described because there is no alternative. This is not the case. In particular, concerning the desirability of an asset class with a particular return pattern, a frequent practice now is to run an efficient frontier with and without the asset class. (This practice is subject to the essential caveat that future returns are not necessarily like those of the past, but the CAPM adjustment is subject to this same caveat.) The comparison of frontiers with and without the asset class avoids Assumptions 4 and 4′ and Conclusion 2.

Concerning the choice between two projects, I previously considered their effect on a company's stock price under the assumption that the stock appears in some but not all segments of investors' efficient frontiers (Markowitz 1990). The resulting computation is similar to that of the CAPM but involves only investors who own the company's stock. In other words, the calculation takes into account the company's clientele. For estimating the effects of investment policy in a dynamic world with mean–variance investors holding different beliefs and real-world constraints, Jacobs et al. (2004) proposed detailed, asynchronous simulation. Potentially, the simulated market could also include investors other than those with mean–variance objectives.[3] In sum, the position that "there is no alternative" to the CAPM for risk-adjustment calculations was never completely true and is certainly not true now.

So What #4. The implications of the CAPM are taught to MBA students and CFA charterholders. The lack of realism in A4 and A4′ is rarely pointed out, and the consequences of replacing these assumptions with more realistic assumptions are rarely (if ever) discussed. Worse, often the distinction between the CAPM and mean–variance analysis is confused. Not only do some say or suggest that if investors use mean–variance analysis, C1 and C2 will follow; some say or suggest that if an investor uses mean–variance analysis, C2 should be assumed to hold among inputs.

Despite its drawbacks as illustrated here, the CAPM should be taught. It is like studying the motion of objects on Earth under the assumption that the Earth has no air. The calculations and results are much simpler if this assumption is made. But at some point, the obvious fact that, on Earth, cannonballs and feathers do not fall at the same rate should be noted and explained to some extent. Similarly, at some point, the finance student should be shown the effect of replacing A4 or A4′ with more realistic constraints and the "so what" should be explained.

About 30 years ago, Fama (1976), in Chapter 8, explained the main points of the present article: that A4 or A4′ are not realistic and that if more realistic assumptions are substituted, C1 and C2 no longer follow. The two principal differences between Fama's presentation then and the current presentation are (1) my use of certain ("portfolio space") diagrams to illustrate the source and possible extent of the market portfolio inefficiency and (2) our respective conclusions concerning "so what." Fama's conclusion at the time was that what one could say about models with more realistic versions of A4 or A4′ was that they "fall substantially short of interesting and testable propositions about the nature of capital market equilibrium. For such propositions, we have to rely on" the CAPM (p. 305). My own conclusion is that it is time to move on.

Conclusion

The CAPM is a thing of beauty. Thanks to one or another counterfactual assumption, it achieves clean and simple conclusions. Sharpe did not claim that investors can, in fact, borrow all they want at the risk-free rate. Rather, he argued:

> In order to derive conditions for equilibrium in the capital market we invoke two assumptions. First, we assume a common pure rate of interest, with all investors able to borrow [without limit] or lend funds on equal terms. Second, we assume homogeneity of investor expectations. Needless to say, these are highly restrictive and undoubtedly unrealistic assumptions. However, since the proper test of a theory is not the realism of its assumptions but the acceptability of its implications, and since these assumptions imply equilibrium conditions which form a major part of classical financial doctrine, it is far from clear that this formulation should be rejected—especially in view of the dearth of alternative models leading to similar results. (pp. 433–434)

Now, 40 years later, in the face of the empirical problems with the implications of the model, we should be cognizant of the consequences of varying its convenient but unrealistic assumptions. In particular, we should be cognizant of what more realistic assumptions concerning investment constraints imply about how we should invest, value assets, and adjust for risk.

Appendix A. Finding a Probability Distribution for a Given Efficient Set

To construct a probability distribution of coconut production whose means, variances, and covariances imply a specific three-security efficient set, you may proceed as follows. The simplest distribution to construct with the requisite efficient set is a finite population with S equally likely sample points, $s = 1, \ldots, S$, with r_i^s as the return on security i if sample point s occurs. The procedure is as follows:

First, use the procedure described in Chapter 11 of Markowitz (1987) or Markowitz and Todd to produce an expected return vector, $\boldsymbol{\mu}$, and a covariance matrix, $\mathbf{C}$, that gives rise to the specified efficient set. (There are always many $\boldsymbol{\mu}$'s and $\mathbf{C}$'s that will serve. Start with any.) This step is not necessary if a $\boldsymbol{\mu}$ and a $\mathbf{C}$ are already given, as in the example in the text.

Second, by using a program that finds the eigenvalues and eigenroots of $\mathbf{C}$, you can find a matrix $\mathbf{B}$ such that $\mathbf{C} = \mathbf{B}'\mathbf{B}$.[4]

Third, let $\mathbf{R}^a$ be a matrix containing a finite sample space for three random variables with 0 mean and covariance matrix $\mathbf{I}$. For example,

$$\mathbf{R}^a = \frac{\sqrt{2}}{4}\begin{pmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

$$= r_{i,j}^a,$$

where $r_{i,j}^a$ is the value of the ith random variable in state (sample point) j.

Then, $\mathbf{R}^b = \mathbf{B}\mathbf{R}^a$ is the matrix of a sample space of three random variables with 0 mean and covariance matrix $\mathbf{C}$. And $\mathbf{R}^c = (r_{i,j}^b + \mu_i)$ has covariance $\mathbf{C}$ and expected return μ. If $\mathbf{R}^c$ has any negative entries and if k is the magnitude of the largest in magnitude negative $r_{i,j}^c$, then $\mathbf{R}^d = (r_{i,j}^c + k)$ is the matrix of a sample space of hypothetical coconut production with nonnegative output and with the specified efficient set.

Notes

1. Some conclusions remain unchanged if we assume heterogeneous rather than homogeneous beliefs; other conclusions apply to average predictions rather than unique predictions.
2. A3 asserts that the market is strong-form efficient in the Fama (1970) taxonomy. Thus, what I will show is that, even if the market is strong-form efficient, the market portfolio is not necessarily a mean–variance efficient portfolio.
3. Simulation analysis as presented by Jacobs et al. (2004) would hardly have been feasible in 1964 when Sharpe presented the CAPM. Computer and software development since that time makes such simulation quite manageable.
4. For example, see Franklin (2000). The formula is a corollary of Section 4.7, Theorem 3. Also, see Section 7.3, Equation 21 for an alternative factorization of **C**.

References

Fama, Eugene F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance*, vol. 25, no. 2 (May):383–417.

———. 1976. *Foundations of Finance: Portfolio Decisions and Securities Prices.* New York: Basic Books.

Franklin, Joel N. 2000. *Matrix Theory*. Mineola, NY: Dover Publications.

Jacobs, Bruce I., Kenneth N. Levy, and Harry M. Markowitz. 2004. "Financial Market Simulation." *Journal of Portfolio Management* (30th Anniversary):142–152.

———. Forthcoming. "Portfolio Optimization with Factors, Scenarios, and Realistic Short Positions." *Operations Research*.

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

———. 1990. "Risk Adjustment." *Journal of Accounting, Auditing and Finance*, vol. 5 (Winter/Spring):213–225.

Markowitz, Harry M., and Peter Todd. 2000. *Mean–Variance Analysis in Portfolio Choice and Capital Markets* (revised reissue with chapter by Peter Todd). New Hope, PA: Frank J. Fabozzi Associates.

Roll, Richard. 1977. "A Critique of the Asset Pricing Theory's Tests, Part I: On Past and Potential Testability of the Theory." *Journal of Financial Economics*, vol. 4, no. 2 (March):129–176.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance*, vol. 14, no. 3 (September):425–441.

Tobin, J. 1958. "Liquidity Preference as Behavior towards Risk." *Review of Economic Studies*, vol. 25, no. 2 (February):65–86.

