Topics in Applied Investment Management: *From a Bayesian Viewpoint*

Harry M. Markowitz

Harry M. Markowitz
is a professor of finance at the Rady School of Management in La Jolla, CA.
harryhmm@aol.com

A few months ago I promised John Guerard that I would write an introduction to the articles for this special issue of *The Journal of Investing*. Now that I have completed this task, I see that my remarks are more like discussant comments than an introduction and could equally well be read after reading the articles. We should have anticipated this outcome, because I hold strong views on at least some of the issues in these articles.

The articles assembled in this issue seek to extend the frontiers of applied quantitative investment management. In particular, Tsuchida, Zhou, and Rachev explore various matters, including doing portfolio selection other than in what they call the "Markowitz way." The other articles mostly seek to do "the Markowitz way" (i.e., mean–variance analysis) better. Specifically, the article by Saxena and Stubbs may be viewed as trying to get the anticipated, ex ante efficient frontier closer to the eventually realized, ex post frontier by tailoring covariance estimates to the expected return model and constraints used in the analysis. Wormald and van der Merwe illustrate the applicability of new computational techniques by seeking "via shrinkage techniques for a better estimate $V_{est} \neq V_{sample}$ for the covariance matrix to be used within an optimization." Clark and Kenyon show how to find efficient portfolios in a many-dimensional space, including mean, variance, and other desiderata. Computational difficulties arise from the non-convexity of some objectives. The two articles by Guerard et al. concern a model for estimating expected returns from fundamental factors, analyst forecasts, and momentum.

Not surprisingly, I am more supportive of the articles that seek to perfect or extend mean–variance analysis than those that question it. However, my views on all the articles are very much conditioned by my basic philosophy concerning portfolio selection, which is contained in Chapter 6 and the last four chapters of Markowitz [1959]. A large number of subsequent research articles, by me and others, have followed up along the same lines, but among the population of financial theorists and practitioners as a whole, those who are familiar with my fundamental assumptions form—to a first approximation—a set of measure zero. Consequently, I will first summarize my fundamental assumptions and related work and then apply this viewpoint to the articles of this special issue.

MARKOWITZ'S FUNDAMENTAL ASSUMPTIONS

Markowitz [1959] justifies mean–variance analysis by relating it to the theory of rational decision making over time and under uncertainty, as formulated by von Neumann and Morgenstern [1944]; Savage [1954] and Bellman [1957]. The fundamental assumptions

of the book appear in Part 4, Chapters 10–13. Specifically, Chapter 10 deals with single-period decision making with known odds. It echoes the view that, in this case, the rational decision maker (RDM) may be assumed to follow certain axioms from which follow the expected utility maxim. (Here I mean the "expected utility maxim" in the sense of Bernoulli [1954] and von Neumann and Morgenstern [1944], in which a utility is attached to each possible "outcome," e.g., to each possible return rather than to a combination of risk and expected return as is sometimes meant.) Chapter 11 of my 1959 book considers many-period games, still with known odds. It shows that essentially the same set of axioms as that in Chapter 10 implies that an RDM would maximize expected utility for the game as a whole. It notes that this indicates that the RDM would also maximize the expected values of a sequence of *single-period* utility functions, each using a Bellman "derived" utility function. Chapter 12 considers single- or multiple-period decision making with unknown odds. Taking off from Savage's work, it adds a "sure thing" principle to the axioms of Chapters 10 and 11 and concludes that when odds are unknown, the RDM maximizes expected utility using "probability beliefs" when objective probabilities are not known. These probability beliefs shift according to Bayes rule as evidence accumulates.

Chapter 13 applies the conclusions of Chapters 10–12 to the portfolio selection problem. It extends an observation, already made in Chapter 6 for the logarithm utility function, that if the probability distribution of a portfolio's returns is not "too spread out," then a function of its mean and variance closely approximates its expected utility. As Chapter 11 explains, the utility function being approximated is Bellman's derived utility function. Typically, it is assumed that a consumption/investment game has only one state variable, namely, end-of-period portfolio value. But Chapter 13 also considers how a mean–variance analysis can approximate expected utility when the utility function contains other state variables, such as other random sources of return (not subject to change within the portfolio analysis, but perhaps correlated with portfolio return) or state variables predictive of future return distributions.

The reason these fundamental assumptions are treated in the back rather than the front of the book is that I feared that if I had started the book with an axiomatic treatment of the theory of rational decision making under uncertainty, no one involved with managing money would read it.

Subsequent research. Chapters 6 and 13 of Markowitz [1959] illustrate empirically, with a few utility functions and return distributions, that, historically, expected utility of return can be fairly closely approximated by

$$EU(R) \cong U(0) + U'(0)\, E + 0.5U''(0)(V + E^2) \quad (1a)$$

or, usually better,

$$EU(R) \cong U(E) + 0.5U''(E)V \quad (1b)$$

where $U(X)$, $U'(X)$ and $U''(X)$ represent the utility function and its first two derivatives evaluated at portfolio return $R_p = X$, and E is either the expectation operator or (standing alone) the expected return of a portfolio. The right side of Equation (1a) is based on a second-degree Taylor approximation around zero return; that in Equation (1b) is centered at return equal expected return E. Young and Trent [1969] found similar results for the logarithmic utility function for a much larger sample of historical return distributions.

Levy and Markowitz [1979] compare average utility versus mean–variance approximations for various utility functions and historical distributions. As in Markowitz [1959] and Young and Trent [1969], Levy and Markowitz [1979] assume that distributions of returns are the same as various historical distributions. Distributions used include annual returns on 149 investment company portfolios, annual and monthly returns on 97 individual stocks, and annual returns on randomly drawn five- and six-security portfolios. The utility functions analyzed were

$\log_e(1 + R)$,

$(1 + R)^a$ for $a = 0.1, 0.3, 0.5, 0.7$ and 0.9, and

$-\exp[-b(1 + R)]$ for $b = 0.1, 0.5, 1.0, 3.0, 5.0$ and 10.0

For $U = \log_e(1 + R)$, the correlation between $E(U)$ and a mean–variance approximation to it, over the 149 distributions of annual returns on investment companies, was $\rho = 0.997$. That for $U = (1 + R)^{1/2}$ was $\rho = 0.999$. For all utility functions except $-\exp[-b(1 + R)]$ for $b = 3.0$, 5.0, and 10.0, the correlation between $E(U)$ and $f(E,V)$ was at least 0.997. In most cases, it was 0.999.

For $-\exp[-b(1 + R)]$, $b = 3.0, 5.0,$ and 10.0, the correlations were, respectively, 0.949, 0.855, and 0.447. Concerning the "misbehaved" $-\exp[-10(1 + R)]$, Levy and Markowitz argue that while all utility functions satisfy

the von Neumann and Morgenstern axioms, not all will appeal to many (or any) investors. In particular,

$$U = -\exp\left[-10(1 + R)\right] \tag{2}$$

exhibits certain strange preferences. For example, an investor with this utility function would prefer a 10% return with certainty to a 50–50 chance of zero return (no gain, no loss) versus a blank check (e.g., a gain of 1,000,000,000%). Levy and Markowitz refer to this utility function as being pathologically risk averse. Markowitz, Reid, and Tew [1994] refer to it as having an implausibly low "value of a blank check" and present the results of a survey suggesting a range of plausible values.

Correlations between E(U) and its mean–variance approximation were smaller over the set of annual returns on 97 individual stocks. For example, for $U = \log_e(1+R)$, the correlation was 0.880 for the annual stock returns, compared with 0.997 for the annual investment company returns. (The mean–variance criteria is to be applied to the portfolio as a whole, but single-stock portfolios were used as examples of "portfolios" with greater variability.) When the holding period was reduced, the correlations increased, as expected, since monthly return distributions are more compact than those of annual returns. For example, for monthly stock returns the correlation for $U = \log_e(1 + R)$ was 0.995. Also, a bit of diversification helps a lot. For annual returns on 19 small (five- or six-security) non-overlapping portfolios drawn at random from the 97 securities, the correlation for $U = \log_e(1 + R)$ was 0.998. Similar effects (of going from investment companies to individual stocks, from annual stock returns to monthly stock returns, and from annual stock returns to annual returns on five- or six-stock portfolios) were found for the other utility functions.

Thus, for most utility functions considered, mean–variance approximations did quite well, especially for annual returns on diversified portfolios, and even for monthly returns of undiversified portfolios.

Subsequent studies include Dexter, Yu, and Ziemba [1980], Pulley [1981, 1983], Kroll, Levy, and Markowitz [1984]; Ederington [1995], and Hlawitschka [1994]. Dexter, Yu, and Ziemba [1980] and Simaan [1993] each assume a particular (different in the two cases) functional form for the distribution of security returns, fit its parameters to historical data, and solve analytically for the loss due to using a mean–variance approximation. Pulley [1983] assumes particular functional forms, generates synthetic "histories" of joint returns from these forms, and proceeds as if the samples were the joint distributions. Ederington [1995] evaluates mean–variance approximations in terms of "10,000 simulated years" by drawing four quarters each at random from quarterly returns on 130 mutual funds. He argues that these 10,000 synthetic years for each mutual fund provides more extreme cases and, therefore, a better test of mean–variance approximations than the relatively few years of actual history used by Young and Trent [1969]; Levy and Markowitz [1979], and others.

It is widely held that the inclusion of puts or calls in a portfolio makes mean–variance analysis inapplicable, since such instruments are "non-normal" and not linearly related to the remainder of the portfolio. But Hlawitschka [1994] considers portfolios of calls (with a 10% allocation to Treasury securities to avoid default) and finds that mean–variance approximations do quite well with portfolios of 10 calls each.

The conclusions of the above studies are generally supportive of mean–variance approximations. However, Grauer [1986] illustrates that in the case of highly leveraged portfolios—as permitted by the capital asset pricing model (CAPM) constraint sets—mean–variance approximations may do poorly. Loistl [1976] is highly critical of mean–variance as an approximation to expected utility, but this is because he erroneously treats a 30% gain or loss as $R = \pm 30$ rather than $R = \pm 0.3$.

Markowitz [2012] compares six methods of approximating expected log (or equivalently, geometric mean) and finds that three of them do quite well (including that in Equation (1b) but not that in Equation (1a) even in databases with extreme moves, such as the real returns to the German and Japanese stock markets during the twentieth century. This ability of a function of mean and variance to approximate expected log utility in the face of an occasional large downward movement is especially significant if one is concerned with extreme values, since $\log(1 + R) \to -\infty$ as $R \to -1$.

Markowitz and van Dijk [2003] consider the loss of utility when a "mean–variance surrogate" is used in place of the derived utility function in certain dynamic investment games with trading costs and changing probability distributions. They report little loss in expected utility compared with an optimum strategy. Their method can be scaled up to apply to much larger investment games— ones that cannot be solved by current dynamic optimization techniques—such as games with too many state

variables for dynamic programming or too many time periods for stochastic programming. Kritzman, Myrgren, and Page [2009] report on the successful use of the Markowitz–van Dijk approximation in rebalancing client portfolios at State Street Bank. The significance of this line of research for the present discussion is that it further demonstrates the ability of functions of mean and variance to approximate real-world derived utility functions.

Von Neumann and Morgenstern [1944], Savage [1954], and their followers have convinced many that rational action in the face of risk or uncertainty involves maximizing expected utility. The calculations cited above indicate that for certain single-period utility functions and historic return distributions, little is lost by using mean–variance approximations to expected utility. The disadvantage in using such an approximation is that it does not provide the precisely optimal solution. Its advantages are that it typically takes less computer time to optimize; uses relationships that hold independent of the form of the distribution, as long as the latter has finite first and second moments; often has fewer parameters to estimate; and last but not least, can be used without ascertaining the investor's actual utility function. By carefully picking a point from a mean–variance efficient frontier, the investor can approximately maximize expected utility—*even without anyone knowing his or her utility function!*

APPLICATIONS TO TOPICS IN APPLIED INVESTMENT MANAGEMENT

Re: Mean-Expected Tail Loss (ETL) Portfolio Selection. To my knowledge, analyses of the ability of risk–return analysis to approximate the expected values of concave (risk-averse) utility functions have not been performed with measures of risk other than variance. It seems to me that proponents of alternate risk measures (such as ETL, used by Tsuchida, Zhou, and Rachev in their article in this issue) should either explain why they do not accept the expected utility maxim or show that their risk measure approximates expected utility better than variance for relevant utility functions and probability distributions. The former charge, to refute expected utility, carries a heavy burden, because much of the modern theory of rational decision making under risk and uncertainty is based on it. On the other hand, the exercise of seeing whether some other risk measure can do better than variance in such an approximation is not a particularly

burdensome task, but it is one that the proponents of alternate risk measures never offer.

Re: Factor Alignment Problems. The variances and covariances that a Bayesian decision maker would use if he or she sought mean–variance efficiency would concern the deviations of actual returns from the RDM's expected returns. In particular, the variance of an RDM's estimate for a security or portfolio return R can be divided into two parts:

$$
\begin{aligned}
V(R) &= E(R - \bar{E})^2 \\
&= E((R - E_h) + (E_h - \bar{E}))^2 \\
&= V(R - E_h) + V(E_h - \bar{E}) \quad (3)
\end{aligned}
$$

since $E(R - E_h)(E_h - \bar{E}) = 0$, where E_h is the expected value of R if hypothesis h is true, and $\bar{E}$ is the expected value of E_h averaged over all hypotheses, weighted by the current probability (or probability density) that the RDM attaches to hypothesis h.

The first term on the right side of the last line of Equation (3) is essentially the average variance; the second term is the variance of the average. Thus, the estimate called for here is *not* what the decision maker considers the *most likely* hypothesis concerning return variance. It also involves the RDM's uncertainty in the estimate of E_h.

The work by Saxena and Stubbs on factor alignment problems is concerned with the dependence of proper covariance estimation on the expected return methodology used. Any such methodology may be viewed as a hypothesis h used to estimate E_h. In other words, their focus is on the value of the covariance structure behind the $V(R - E_h)$ term—*for a given h*—and how this varies from one h to another. The $V(R - E_h)$ term in Equation (3) is an average of such, to which should be added $V(E_h - \bar{E})$.

One thing especially surprising about the Saxena and Stubbs results is the dependence of actual ex post covariance matrices on the constraints used—specifically, on the varying degree of "bindingness" of a constraint, as indicated by its varying shadow price. Perhaps the way to interpret this result is to admit that we can aspire to be only imperfect imitations of an RDM; and since portfolio optimization in practice is a process involving constraints as well as expected return estimates, the variability in estimation errors made by such a process depends on the variability of the bindingness of the constraints used as well as on its method of expected return estimation.

Re: Constrained Optimization for Portfolio Construction. Wormald and van der Merwe present "an economically important optimization problem that can be solved efficiently by means of techniques such as second-order cone programming that allow multiple quadratic constraints." The specific "economically important problem" analyzed concerns the better estimation of covariance matrices. This is certainly an important practical goal, but I wonder whether cone programming is any better than an older method for the problem posed.

As in the case considered in their article, let $Q_1(X)$ and $Q_2(X)$ be two positive semi-definite quadratic functions of portfolio X. Let $E(X)$ be the portfolio's expected return. X is to be chosen subject to the same kinds of linear equality and inequality constraints (such as a turnover constraint and upper and lower bounds on individual positions) as permitted by the general portfolio selection model. (Markowitz [1959], Chapter 8 and Appendix A.)

One way to obtain efficient points in (E, Q_1, Q_2) space is to fix Q_2 and trace out the efficient (E, Q_1) combinations with Q_2 thus fixed, repeating this for other values of Q_2. Wormald and van der Merwe use this procedure for a particular Q_2, constrained to be less than or equal to 1.5 %, 2.3%, and unbounded. (They don't explain why they chooe 2.3% as a bound—or 1.5% for that matter.)

Another way to generate (E, Q_1, Q_2) efficient portfolios would be to assume a certain rate of substitution between Q_1 and Q_2:

$$Q = \alpha Q_1 + (1-\alpha)Q_2 \qquad (4)$$

for $\alpha \in [0,1]$; trace out the (E, Q) efficient frontier for a given α; then repeat for different values of α.

Since Q_1 and Q_2 are positive semi-definite, so is Q. Thus, the critical line algorithm (CLA) (Markowitz [1959]) can be used for tracing out an (E, Q) frontier. Niedermayer and Niedermayer [2010, p. 383] report: "In order to compare [a new portfolio selection algorithm] with the CLA, we implement a numerically enhanced version of the algorithm in Fortran 90. We show that this algorithm outperforms the algorithm in Steuer et al. [2006] by a factor of almost 10,000 (for 2,000 assets) and standard software packages by even more." Specifically, they show in their Table 12.1 that the CLA computes the 2000-security efficient frontier in 0.78 seconds, whereas its closest competitor, 6,300 seconds. If the "second-order cone programming" proposed by Wormald and van der Merwe is little, if any, faster than the competitors of

the CLA tested by Niedermayer and Niedermayer, then "strips" from the (E, Q_1, Q_2) frontier could be traced out by the CLA for

$$\alpha = 0.0, \ 0.1, \ \ldots, \ 1.0$$

or even for

$$\alpha = 0.0, \ 0.01, \ 0.02, \ \ldots, \ 1.0$$

faster than two or three strips could be computed for fixed Q_2 using second-order cone programming. The (E, Q_1, Q_2) combinations on the strips computed by the CLA could be displayed in a 3-D graph. I believe this would be more valuable in practice than the results for two or three values of Q_2.

Re: Multiobjective Evolutionary Algorithms (MOEAs). Clark and Kenyon include the possibility of non-convex constraints in tracing out efficient frontiers. An example is an upper bound on the number of securities held. This can be formulated as a "mixed integer quadratic programming problem," which is much harder to solve than minimizing a positive semidefinite quadratic subject to a convex constraint set. Algorithms in the "mathematical programming" tradition—especially the integer programming tradition—are available (commercially and otherwise) for solving such problems. (An internet search on "mixed integer quadratic programming" shows the latest proposals and offerings in this field.) It would be interesting to see a comparison between the computer times required by Clark and Kenyon's MOEAs and the more traditional methods for this type of problem.

Re: The United States Expected Return (USER) Model. The two articles by Guerard et al. continue a line of research that Guerard has pursued for many years. This research seeks good expected return estimates for individual securities, altering these expectations from period to period (e.g., month-to-month in the current model) as company attributes and investing fashions change.

Guerard's work may be characterized as pragmatic and conscientious. It is pragmatic in the sense that it uses "what works." Specifically, his models use variables that have been predictive of expected returns, on average, over some relatively long period of time. Each period, he refits the regression coefficients of security returns against these variables to reflect what has worked best recently. His work is conscientious, for example, in his use of robust regression, data mining corrections, and the frequent review of alternatives.

My one specific recommendation concerning Guerard's methodology, made some years ago, concerned data mining corrections, as described in Markowitz and Xu [1994]. The shrinkage-of-estimates recommended therein (based on an "empirical *Bayesian*" approach) has been incorporated into Guerard's process. There inevitably remains the generic caveat that a theory or process that has worked until today may not work tomorrow. The most notable example of this in quantitative finance was the failure of the relationships on which Long Term Capital Management was based. This "failure of relationships" was only temporary, since they depended on arbitrages that had to work—eventually. But with LTCM's highly leveraged portfolio, "eventually" was not soon enough. The moral of this story is: Don't highly leverage on the basis of *any* model or process.

Much less disastrous, but more immediately relevant, was the inability of the model in Bloch et al. [1992] to keep up, in the late 1990s, with a rising Japanese stock market that seemed little concerned with "value." Since then, momentum has been added to the list of variables of the USER model, to good effect, as reported in the two articles by Guerard et al. Nevertheless, there is always the chance that the current model too will stop working, at least for a while.

That said, I must congratulate John Guerard for putting together a collection of articles of great theoretical interest, including models and methods for which strong evidence is presented that they will *probably* continue to work in practice, most of the time, with an occasional tweak or two. That is as much as one can ask for in this uncertain world.

REFERENCES

Bellman, R.E. *Dynamic Programming.* Princeton University Press, 1957.

Bernoulli, D. "Specimen theoriae novae de mensura sortis. Exposition of a New Theory on the Measurement of Risk" (English translation by Louise Sommer). *Econometrica*, 22 (1954), pp. 23-26 (originally published in 1738. Comm. Acad. Sci. Imp. Petropolitanae, 5, pp. 175-192).

Bloch, M., J. Guerard, H. Markowitz, P. Todd, and G. Xu. "A Comparison of Some Aspects of the U.S. and Japanese Equity Markets." *Japan and the World Economy*, 5 (1993), pp. 3-26.

Dexter, A.S., J.N.W. Yu, and W.T. Ziemba. "Portfolio Selection in a Lognormal Market When the Investor Has a Power Utility Function: Computational Results." In *Stochastic Programming*, edited by M.A.H. Dempster. New York: Academic Press, 1980, pp. 507-523.

Ederington, L.H. "Mean–Variance as an Approximation to Expected Utility Maximization; Semi-Ex-Ante Results." In *Advances in Financial Economics*, Vol. 1, edited by Mark Hirschey and Wayne Marr, 1995.

Grauer, R.R. "Normality, Solvency, and Portfolio Choice." *Journal of Financial and Quantitative Analysis*, Vol. 21, No. 3 (1986), pp. 265-278.

Hlawitschka, W. "The Empirical Nature of Taylor-Series Approximations to Expected Utility." *The American Economic Review*, Vol. 84, No. 3 (1994), pp. 713-719.

Kritzman, M., S. Myrgren, and S. Page. "Optimal Rebalancing: A Scalable Solution." *Journal of Investment Management*, Vol. 7, No. 1 (2009), pp. 9-19.

Kroll, Y., H. Levy, and H.M. Markowitz. "Mean–Variance Versus Direct Utility Maximization." *Journal of Finance*, Vol. 39, No. 1 (1984), pp. 47-61.

Levy, H., and H.M. Markowitz. "Approximating Expected Utility by a Function of Mean and Variance." *American Economic Review*, Vol. 69, No. 3 (1979), pp. 308-317.

Loistl, O. "The Erroneous Approximation of Expected Utility by Means of a Taylor's Series Expansion: Analytic and Computational Results." *The American Economic Review*, Vol. 66, No. 5 (1976), pp. 904-910.

Markowitz, H.M. *Portfolio Selection: Efficient Diversification of Investments.* Wiley, 1959; Yale University Press, 1970; Basil Blackwell, 1991.

—— "Mean–Variance Approximations to the Geometric Mean." *Annals of Financial Economics*, forthcoming, 2012.

Markowitz, H.M., and E. van Dijk. "Single-Period Mean–Variance Analysis in a Changing World." *Financial Analysts Journal*, Vol. 59, No. 2 (2003), pp. 30-44.

Markowitz, H.M., and G.L. Xu. "Data Mining Corrections: Simple and Plausible." *The Journal of Portfolio Management*, Vol. 21 (Fall 1994), pp. 60-69.

Markowitz, H.M., D.W. Reid, and B.V. Tew. "The Value of a Blank Check." *The Journal of Portfolio Management*, Vol. 20, No. 4 (1994), pp. 82-91.

Niedermayer, A., and D. Niedermayer. "Applying Markowitz's Critical Line Algorithm." In *Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques*, edited by John B. Guerard, Jr. New York: Springer, 2010, pp. 383-400.

Pulley, L.M. "A General Mean–Variance Approximation to Expected Utility for Short Holding Periods." *Journal of Financial and Quantitative Analysis*, 16 (1981), pp. 361-373.

——. "Mean–Variance Approximations to Expected Logarithmic Utility." *Operations Research*, Vol. 31, No. 4 (1983), pp. 685-696.

Savage, L.J. *The Foundations of Statistics*. John Wiley & Sons, 1954. Second revised ed., Dover Publications, 1972.

Steuer, R.E., Y. Qi, and M. Hirschberger. "Portfolio Optimization: New Capabilities and Future Methods." *Zeitschrift für BWL*, Vol. 76, No. 2 (2006), pp. 199-219.

Simaan, Y. "What is the Opportunity Cost of Mean-Variance Investment Strategies?" *Management Science*, Vol. 39, No. 5 (1993), pp. 578-587.

Von Neumann, J., and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, 1944. 3rd ed., 1953.

Young, W.E., and R.H. Trent. "Geometric Mean Approximation of Individual Security and Portfolio Performance." *Journal of Financial Quantitative Analysis*, 4 (1969), pp. 179-99.