Yale ICF Working Paper No. 09-03

A DISCRETIONARY WEALTH APPROACH TO INVESTMENT POLICY

Jarrod W. Wilcox
Wilcox Investment

Frank Fabozzi
Yale School of Management

A DISCRETIONARY WEALTH APPROACH TO INVESTMENT POLICY

Jarrod W. Wilcox

Wilcox Investment

Frank J. Fabozzi

Yale School of Management

A DISCRETIONARY WEALTH APPROACH TO INVESTMENT POLICY

Despite portfolio construction based on expected utility theory and Markowitz mean-variance optimization having been the foundation of financial economic theory for more than 50 years, its practical application by financial advisors has been limited. Particularly troubling are the lack of a normative risk-aversion parameter customized to individual investor circumstances and the need for extensive constraints to produce practically acceptable results. We propose a comprehensive conceptual framework for better investment policy. To begin, we develop investor circumstance-contingent risk aversion for use in portfolio construction, taking into account higher moments of return only as needed. We recursively maximize expected logarithmic return on what we define as "discretionary wealth" to generate many-period maximization of median wealth without violating interim shortfall points. Then we extend this basic framework based on point estimates to fully Bayesian logic. This offers not only improved decision inputs but the advantage of deriving the probability distribution of the objective as a function of portfolio weights before selecting the best portfolio.   Implications are discussed for a wide array of practical issues: investor leverage, longevity risk, higher return moments, dynamic hedging, life-cycle investing, performance measures, and robust portfolio construction. Hypotheses for implied market structure and pricing are also generated.

A DISCRETIONARY WEALTH APPROACH TO INVESTMENT POLICY

1. INTRODUCTION

New ideas are more readily accepted if they are incremental to well-established paradigms. And yet, as John Maynard Keynes wrote in the preface to *The General Theory of Employment, Interest and Money*, "The difficulty lies, not in the new ideas, but in escaping from the old ones…." It has been more than a half century since the publication of von Neumann and Morgenstern's *Game Theory and Economic Behavior* introducing risk into utility curves and of Harry Markowitz's 1952 article on an improved approach to optimizing portfolio construction using estimates of risk aversion together with return means, variances, and covariances. The von Neumann-Morgenstern expected utility theory has been taken as based on the subjective preference of the individual, with little guidance on what makes a good risk and return tradeoff. Markowitz mean-variance optimization limits itself to variance as a measure of risk, and again omits guidance as to the tradeoff one should make between expected return and risk. Both these paradigms counterfactually assume we know their input parameters with certainty.

Continuing research on normative models is needed. Possibly no one is better aware of this than Markowitz, as seen not only in his early pondering of long-run growth criteria (Markowitz, 1976), but also in later work exploring the use of the semi-variance, as well as in testing improvements through resampling and Bayesian methods (Markowitz and Usmen, 2003). However, it is the simplicity of his original version that still has pervasive influence. Its reduction of risk to a linear function of variance in investment returns is not well-suited to an investor confronting increased use of leverage and derivative securities with option-like characteristics. Even for plain vanilla securities, the non-linear structure of the portfolio optimization problem conflicts with the assumption of the adequacy of point estimates as decision inputs, requiring the application of *ad hoc* constraints to produce usable results in many practical cases, particularly in long-short portfolios. Finally, a subjectively chosen risk-aversion coefficient is ill-suited to the household investor. This lack of specification for appropriate risk aversion is especially troubling. As Campbell (2006) noted in his Presidential address to the American Finance Association, "A fundamental issue that confronts the normative literature is how to specify the household utility function…. Until some consensus is reached, normative household finance should emphasize results that are robust to alternative specifications of household utility." We take Campbell's statement as an appropriate challenge and a stimulus to innovation. Utility, which has trouble gracefully representing problems with path dependency, bears too heavy a burden as the primary tool for dealing with multiple periods, especially if appropriate risk aversion is contingent on prior outcomes.

We are far from alone in recognizing the need for change, and others have made important advances. However, research results beyond the more established paradigms of finance have had surprisingly little success in diffusing better investing into general practice. For example, stochastic differential equations (Merton, 1990) have yielded important insights, but the resulting support for models integrating consumption and spending decisions, as in the model of Dybvig (1995), has not discernibly influenced financial planners. The recognition of

3

the need for a more informed input than subjective estimates of risk aversion and for a better connection with interim shortfall avoidance have contributed to applications of the "safety-first" concept early illustrated by Roy (1952). Stutzer (2000, 2003) has explored a more graceful form of safety-first by characterizing portfolios with the fastest rate of decline in probability that a time-averaged return falls below a threshold rate.[1] But this has not, despite its clear practical relevance, been generally accepted for practical use. However, a recent paper by Das and Statman (2009) illustrates an application of the method to optimization of portfolios incorporating structured products with high moment characteristics. Their optimizations are essentially Markowitzian with the addition of a constraint that returns will not be less than some threshold with a probability greater than another threshold. The application of stochastic dynamic programming, surveyed early by Ziemba and Vickson (1975), has shown success in particular applications, and is being used today for some financial planning. But its complexity has proved difficult to scale both in terms of problem size and usage.

Descriptive research in behavioral finance has documented consistent biases away from what is regarded as "rational" processing of information as investment decisions are made. Subsequently more normative models that incorporate some of these investor biases as constraints, such as the use of separate mental accounts rather than integrated optimization, have been suggested, as in Brunel (2006) and Das *et al* (2009). Such models show promise for practical use because they can improve decision-making while conforming to user intuition. We believe, however, that this concession of the necessity of investor biases within normative models is premature, and that even if it should prove desirable, better results can be achieved with a stronger normative foundation.

In this paper, we propose and motivate with indications for applications a normative paradigm for better *investment policy*. This is not an empirical paper, but it does have empirical as well as normative implications, in that investors who follow our suggested investment policy are hypothesized to actually acquire more wealth and have a measurable impact on pricing and clientele structures. This is not a paper focused on tests, but we do provide in Section 3.4 some comparisons of the performance of Markowitz mean-variance optimization with fully Bayesian discretionary wealth methods. Despite implications for practice that could have been used to good advantage in surviving recent unpleasant market surprises and a minimum of mathematical formulae, it is intended as a theoretical contribution.

Our paper illustrates this paradigm within two models. The first model assumes discrete period point estimates for return characteristics arising from outcomes independent through time. Both arithmetic and logarithmic returns are assumed bounded, supporting the application of the Lindeberg central limit theorem (CLT) and mathematical conveniences such as the existence of

---

[1] Stutzer's work is similar in spirit to that of our point-estimate-based model in Section 2, in that he is concerned with the ability to take into account higher return moments and has sought to eliminate the necessity of a purely subjective specification of risk aversion. However, there are many differences. For example, working from a normative premise that the best portfolio is one which has the fastest rate of decay in the probability that time-averaged return will fall below a threshold, he finds that for a normal return distribution, the optimal portfolio is Sharpe ratio maximizing, an assertion which we find true only as a very special case.

Taylor series approximations. Each period, the investor maximizes the expected logarithmic return of *discretionary wealth*, a right-hand balance sheet concept akin to asset-liability surplus. It is calculated as the residual of an augmented balance sheet that includes present values of predicted cash flows to and from the investment portfolio. This process produces normative risk aversion. The investor applies this model recursively so as to approximate maximizing median wealth after many periods. The second model extends the logic of the first so that each quantity, including risk aversion, is treated as a Bayesian probability distribution. This fully Bayesian model greatly improves the robustness of investment policy in cases where conventional Markowitz mean-variance optimization would require additional *ad hoc* modifications to produce acceptable practical results.

We review material on the point-estimate model based on growing discretionary wealth along the lines of Wilcox (2003). This foundation is still largely unfamiliar to academic readers and is necessary to set the stage for the more recent work using a Bayesian approach. In addition, its links with prior literature and the applications of this model have been amplified. The second model based on Bayesian logic is new, although its application to longevity risk is foreshadowed in Wilcox (2008).

The main contributions of this paper are twofold. The first is to improve the application of Markowitz mean-variance optimization, both by providing its missing normative view as to appropriate single-period risk aversion and by restricting its domain to contexts in which it may be most effectively employed — long-only portfolios without excessive investment volatility, with moderate to ample discretionary wealth, and absent strong skewness or kurtosis. The second main contribution is the development of a more comprehensive and robust Bayesian model for investment policy that can work well outside these contexts, as well as accommodate risk in the investor's future consumption plans. To help motivate the adoption of these ideas, numerous applications are briefly introduced, contributing to the effective conceptual integration of a wide span of investment issues.

The reader should note particularly that uncertainty in saving and spending plans is in parallel to uncertainty in the investment returns and is fully comprehended by our Bayesian approach. The implications of the fully Bayesian approach confined to the investment portion are more subtle and do not produce big effects for very simple cases, but are important for cases with long-short positions, high leverage, option positions, or nearly singular covariance matrices. There are also two other innovative sallies in the paper — the implications for market structure and pricing, and the observation that in seeking optimum leverage for a given set of available investment characteristics, we have a powerful way of influencing, though not fully determining, the larger financial planning problem that decides how much to plan on saving and spending.

1.1. Building Blocks of the Paradigm

The first building block for our investment policy paradigm is a better method for the investor to measure his or her current position. This is the discretionary wealth concept. We believe this is a more natural way to represent multi-period path dependency than either "safety-first" rules or constraints on probability of loss. It is similar to asset-liability management in

general, for example by Ibbotson *et al* (2007), with some important differences. Because human capital supports both consumption and investment, we do not treat it. Similarly we do not explicitly treat insurance issues. We treat as implied assets only time-discounted values of predicted future cash flows to the investment portfolio. Implied liabilities apply to foreseeable cash withdrawals from the investment portfolio. Discretionary wealth is the surplus of assets over liabilities including these implied augmentations.

The second building block in the paradigm is a more reliable guide to the investor's efforts to improve his or her position. We take as our objective the maximization of *median discretionary wealth* after many periods. By many periods, we do not necessarily mean long time horizons, but rather that the time horizon may be usefully subdivided into sufficiently many separate points of decision such that policies suitable for an infinite number of periods will be nearly optimal. To achieve the simplicity required to enable broad diffusion among investors, we do not embed transaction costs in our model, except as they may affect expected returns. This omission is no different from that in Markowitz mean-variance optimization. The underlying assumption of this building block is that it is in the investor's best interest to grow discretionary wealth away from the shortfall point of future financial commitments as fast as possible so that he or she will have future choices with regard to increasing consumption-like expenditures. In focusing attention on next-period discretionary wealth rather than on the investment portfolio, our compass tells the investor both *optimal investment risk-aversion* and whether there is a need to go beyond variance to higher statistical moments of return in representing risk. Also, since this immediate objective involves feedback through changing risk-aversion after unexpected discretionary wealth disturbances, it has important self-correcting properties.

It is possible to understand and implement our models without referring to the utility concept. However, those readers who find utility theory indispensable to their thinking may note that our approach is similar to that of Rubinstein (1976) in his description of "generalized logarithmic utility," the generalization being the inclusion of an offsetting quantity to the investor's wealth.

Our pursuit of the multi-period long-run objective of maximizing median wealth after many periods simplifies to engineering maximum expected logarithmic return on discretionary wealth in the current period.[2] Though this linkage from current period objective to multi-period result can be improved if the number of periods is too short to make full use of the CLT, that improvement in precision would be at a cost of model complexity that we defer for more specialized research. Note also that our approach surmounts obstacles raised by Merton and Samuelson (1974) in response to Hakansson's advocacy of growth optimality (1971, 1974), and later by Samuelson (1990). Interpreting their objections to mean that all forms of growth optimality are misguided is too sweeping a conclusion. Growth optimality for discretionary

---

[2] Later in the paper we introduce the single period objective of maximizing expected logarithmic return of the product of the implied leverage ratio (investments to discretionary wealth) and the portfolio return. In the same manner as Markowitz's return objective is tied to a utility objective, our single-period return on discretionary wealth is tied to generalized log optimality as in Rubinstein's paper. Unfortunately, mainstream utility functions such as power utility are subsumed in this formulation only as special cases.

wealth, provided that one has enough periods ahead and that periods are short enough so that self-adaptive risk-aversion prevents intermediate shortfalls, satisfies the requirement for the degree of conservatism needed to protect plans for future withdrawals to support consumption spending. We do not argue that the criterion of maximizing expected log return on discretionary wealth is invariably sufficient. However, we do maintain that it is a considerable improvement on conventional single-period metrics.

The third building block for better investment policy is the recognition by investors that their knowledge is in the form of probability distributions rather than point estimates. The investor should have available the advantage of Bayesian logic when its use makes a material difference. The financial literature, summarized in Fabozzi *et al* (2007), has noted the advantage of Bayesian priors and estimation methods in making the inputs to investment decisions more robust. Going further, however, pioneers in the use of Bayesian methods for utility-based optimization, as in Harvey *et al* (2008), reduce the distortions induced by subjecting point estimates of input probability distribution parameters to non-affine (nonlinear) calculations. These methods and insights need to be widely shared and their consequences appreciated.

Our fourth building block is the incorporation of probability distributions for future cash contributions to, and withdrawals from, the investment portfolio. For example, a household investor desiring to provide for retirement consumption spending does not know how long he or she will live. Ignoring this longevity risk will lead to inaccurate results because information will be lost as to the joint probability of unusually long life and unusually poor investment results. Probability distributions may be assigned to any of the important factors determining discretionary wealth – for example, taxes, changes in employment and family needs, etc. – and consequently assigned to optimal risk aversion parameters. Investment portfolio "optimization" calculated without incorporating these contextual uncertainties that affect appropriate risk aversion can lead to poor investment policy.

1.2 Scope of Paper

There is a strong history of research addressing the more global problem of integrating saving, consumption and investment issues. Early and influential examples include Samuelson (1937), Merton (1990), Grossman and Laroque (1990) and Dybvig (1995). This more global problem seems a critical one for many households and businesses. However, in this paper, we primarily provide tools with which to explore optimal investment policies given saving and spending plans. That is, the investment planning process may be a single step in a more comprehensive iterative process needed to bring consumption plans into effective coordination with investment decisions. Accordingly, we will assume that any revisions of spending or savings plans have already taken place before we address investment questions, understanding that the investment solutions will often be provisional because they may be suboptimal in this larger context.

We can illustrate the difficulty that led us to restrict the scope of the paper in this way with an example based on two assets, a riskless asset with zero returns and a conventional risk-bearing diversified portfolio. Following the analysis to be described later in the paper, we can approximate the optimal ratio of the risk-bearing asset's allocation to discretionary wealth by

setting it equal to the ratio of the risk-bearing asset's expected return to its variance. However, the investor has a choice between adjusting planned consumption so as to affect discretionary wealth and adjusting the risky asset's allocation. A normative model should reflect the costs to that particular investor of a changed consumption plan as compared to a changed investment allocation. To do so in a more objectively determined and comprehensive manner than simply postulating a time-preference parameter ─ which we find problematic ─ seems to be a formidable requirement.

Note again that we are not trying to describe investors as they are, as we would from a behavioral finance viewpoint, but as they ought to be to achieve a specified objective. Though we recognize the applicability of behavioral finance in practice, we do not address the compromises a well-informed investment manager may find advantageous in providing services to a naïve investor, as in Brunel (2006). We also do not confront agency problems. Finally, we apply in our ending examples simple Monte Carlo simulation rather than more advanced Markov chain Monte Carlo methods. This will restrict us to demonstrating Bayesian logic for broad asset allocation rather than providing tools for detailed portfolio management.

1.3 Organization of Paper

We have organized the paper as follows. In Section 2, we construct a framework based on point estimates of discretionary wealth and return parameters to show how to arrive at a contingent risk aversion for use in portfolio construction, period by period, taking into account higher moments of return only as needed. In Section 3, we allow the return decision inputs to be Bayesian probability distributions rather than losing information by compressing them to point estimates. This allows us to use Bayesian logic to construct a probability distribution of logarithmic return on discretionary wealth as a function of portfolio weights before selecting the best portfolio. Having mastered that practice, we can easily extend it by allowing appropriate risk aversion to also be represented as a probability distribution. This takes into account imprecise estimates of future sources and uses of investment funds such as derive from longevity risk.

The evolution in investment policy paradigm we seek will not be undertaken lightly either by researchers or by investors. We have chosen to motivate the approach with short introductions to a multiplicity of practical applications and implications for further research. Consequently no reader interested in a particular application is likely to be satisfied with the depth pursued. The purpose of this paper, however, is to convey the organizing power of a single paradigm for wide use.

Accordingly, Section 2 follows this roadmap: determining appropriate single-period risk aversion, mapping that problem into growing discretionary wealth, and then showing how to provide appropriate risk aversion for Markowitz mean-variance optimization in common circumstances. We then discuss point-estimate methods for investment when skew and kurtosis are important. This is followed by a number of brief indications for application, including the appropriate risk aversions for each higher return moment, managing higher moments through shortening discrete periods between re-allocation, the impact of taxes, integration with consumption decisions, dynamic hedging pitfalls, improved life-cycle investing, and customized

performance measures.   We end Section 2 with a discussion of implications for research in pricing and clientele anomalies.

Section 3 brings Bayesian logic to the model of Section 2, extending it to include probability distributions rather than point estimates for both return parameters and discretionary wealth. After briefly describing what we see as the still underutilized potential for Bayesian analysis in improving decision inputs, we investigate the potential improvements from using fully Bayesian logic in translating decision inputs to probability distributions of our objective function. This potential is fundamental to the often quite nonlinear nature of portfolio optimization. We extend the scope of factors to be considered by including a probability distribution for discretionary wealth based on longevity risk. Finally, we test Markowitz mean-variance optimization relative to a fully Bayesian approach for a limited but instructive set of examples. Robustness is tested using partial knowledge. Algorithm accuracy is tested using fully accurate knowledge of the probability distributions for the inputs. In benign contexts, there is only very slight disadvantage to using the Markowitz algorithm, provided that the risk aversion parameter has been set with cognizance of discretionary wealth. However, the fully Bayesian approach is shown to be meaningfully superior in cases where nonlinearities are important.

Section 4 summarizes our interpretation of the implications of the paper for research and practice.

2. BETTER POINT ESTIMATE INVESTING

It is common in finance to use point estimates as a substitute for probability distributions for representing knowledge to be used in decision-making. For example, one usually calculates the present value of an annuity with a single interest rate. Even though Bayesian methods offer significant advantages under some circumstances, the convenient practicality of point estimate-based calculations for many ordinary investment decisions cannot be ignored. Consequently, we begin our discussion of investing decisions by asking how they might be improved while still using point estimates.

We are far from the first to compare portfolio optimization using Markowitz's approach with some form of utility maximization.[3] With the well-behaved test cases usually employed, and without explicitly allowing for the stronger risk aversion appropriate when discretionary wealth is small compared to the investment portfolio, only small differences are found. This section shows how Markowitz mean-variance optimization can be made into a more normative generalized utility maximizer, but then indicates circumstances when it is likely to need modification to contend with higher return moments.

Those familiar with the historical development of today's modern finance may be aware of the controversy in the early 1970s over whether so-called growth optimal models could be assimilated into utility theory.[4] Mainstream academic finance work thereafter rejected this idea.

---

[3] See, for example, Levy and Markowitz (1979).

[4] See Breiman (1961), Thorp (1971), Hakansson (1971 and 1974), and Merton and Samuelson (1974).

However, a maverick wing of financial researchers has steadfastly taught the value of growth maximization. For examples, particularly the references to "fractional" Kelly strategies (see MacLean and Ziemba, 2006 and Thorp,2006).

The "discretionary wealth" approach put forward here reframes the investment problem to one of growing residual wealth above a shortfall point, as mathematically foreshadowed in Rubinstein (1976), thereby skirting old controversies. Though related to the fractional Kelly approach, we believe it provides an improved mechanism for changing relative risk aversion as total wealth changes its relationship to a shortfall point. Its use of variable risk aversion is also related to the Black and Perold (1992) and Perold and Sharpe (1995) formulation of "constant proportion portfolio insurance" (CPPI). However, unlike CPPI, it takes into account the need to maximize expected growth rate above a shortfall point given investment return characteristics. Finally, as noted earlier, there are "safety-first" cousins to our approach that seek to avoid the subjective assessment of investor risk aversion, for example by focusing on minimizing the probability of shortfall, as in Stutzer (2003).

2.1 Optimal Growth

Begin with a simple example. Consider a game of indefinitely repeated coin flips, with equal probability of "heads" or "tails." The payoff for each outcome is a return of +100% for a head, and −50% for a tail. The expected return each turn is +25% of what we bet, but this does not tell us very much about what we are likely to gain through compounding. Suppose that our objective is to obtain the probability distribution of wealth with the highest *median* after an infinitely large number of turns.

By symmetry across turns, the investment policy decision is to decide what constant fraction of one's wealth to wager at each flip of the coin. If we bet nothing each turn, our median outcome is to end where we began, with an average return per turn of 0%. If we bet our entire wealth each turn, our median outcome is also a 0% return, resulting from an equal number of heads and tails. But if we bet exactly half our wealth each turn, our median outcome will be maximized. The resulting best median outcome implies an average growth rate of only about 6% per turn, illustrating that risk interferes with compounding. We arrive at this best policy by maximizing the expected log return at each turn.

Note particularly that this solution is *myopic*. That is, to achieve the best solution in the long run we can optimize even though knowing only the conditions prevailing in a single turn – an enormous simplification of a multi-period problem.

2.1.1 *Scope*

To what kind of situations might we generalize this reduction in problem complexity? For the purposes of this paper, we confine our attention to return probability distributions each period that are independent of one another, with bounded log returns each period. Then by the Lindeberg CLT (as in Feller, 1968), the probability distribution for the multi-period sum of successive log returns will approach a normal distribution — with its median equal to its mean. Consequently, if we maximize expected log return each period, we maximize the median log return compounded over many periods. Finally, since rank order statistics such as the median are preserved by taking the antilog, we know that we have maximized the median wealth after

10

many periods. Note that in this chain of reasoning the there is no real need for a log-normal distribution; the benefit of the CLT here is the attainment of a symmetric distribution with median equal to mean. Further, there is no need for the single-period return to be either normal or log-normal.

We can afford to limit our assumptions in this way because if, contrary to our assumption, there is a periodic finite probability of a total loss, we know the median long-run wealth is total ruin, and can avoid such a policy in any case. However, in adopting our criterion, we need to keep in mind that it requires modification for dealing with problems where the number of periods contemplated is too small to produce a practically symmetric distribution for the sum of log returns. Then a more accurate procedure would be to consider detailed period-by-period modeling, as with dynamic programming. Recall, though, that the periods can be as short as we are willing to make investment decisions on independently distributed returns given transaction costs, so that an investor in liquid securities facing a deadline a year away might contemplate 12-monthly decisions. (As noted later, this higher frequency of decisions is also an important way to diminish the impact of skew and kurtosis in probability distributions.)

2.1.2 *Taylor Series Representation*

To clarify the determinants of expected log return, we can express log return as a Taylor series expanded around an expected mean arithmetic return as in Booth and Fama (1992), and then take the expectation of each term in the series. This gives a relationship between mean log return and the central moments of the distribution of the ordinary arithmetic return r, as shown in equation (1).

$$\text{Expected } \ln(1+r) \cong \ln(1+E) - \frac{V}{2(1+E)^2} + \frac{SV^{\frac{3}{2}}}{3(1+E)^3} - \frac{KV^2}{4(1+E)^4} + \ldots \qquad (1)$$

where ln is the natural log function, r is the arithmetic return (the conventional, not logarithmic, return measure), E is the mean of r, V is the variance of r, S is the skewness of r, and K is the kurtosis of r (for a normal distribution $K=3$).

Although approximating functions of a random variable as a Taylor series is not uncommon in the scientific literature, one may in general question when it is permissible to express mean log return in this way, since the series still might not converge, or be said to "exist." Generally it seems a satisfactory response to note that practical cases where the Taylor series does not exist or converge are those we would regard as inferior investment strategies. However, for the purpose of simplifying this paper, we assume that both log returns and arithmetic returns are bounded, even when leveraged, and that equation (1) is a valid approximation of expected log return.

Note that if E and V are small, equation (1) will be well approximated by the following equation, often used as a growth rate approximation.

$$\text{Expected } \ln(1+r) \sim E - V/2 \qquad (2)$$

In Section 2.2, we will fundamentally modify these equations to apply to discretionary wealth. But they already provide a window into multi-period investing.

2.1.3 *Taylor Series Implications for the Management of Return Higher Moments*

The difference between equations (1) and (2) gives us insight into the effect of changing the periodicity of investment decisions. For independent, regularly sampled discrete returns of similar distribution, *E* and *V* are approximately linearly related to the length of the time period. This linearity of expected log return with period length dominates when periods are short. However, as periods between decisions lengthen, *V* increases. Then equation (1) shows us that the *relative* effects of the third, fourth, and higher moments, which depend on powers of *V* greater than one, will become increasingly important and the relationship will become materially nonlinear. For example, illiquid investments that force longer time periods between reallocations may make negative skew and fat-tailed return distributions more costly to compounded returns. Moving in the other direction, the ability to shorten the length of individual periods, controlling higher moment effects at a cost of additional transactions, is an important feature of investment policy choice.

A strongly held, but mistaken, assertion found in the finance literature is that the adequacy of Markowitz mean-variance optimization is assured when asset returns follow a normal distribution.[5] Although it is true that a normal distribution is completely determined by its mean and variance, the relationship of variance to logarithmic utility as implied by equation (1) is not the linear one represented by Markowitz's objective function. For example, note from equation (1) that the impact of the fourth return moment is relevant to our objective even for a normal return distribution; it involves a non-zero kurtosis ($K = 3$) times the *square* of the variance. The universal adequacy of Markowitz optimization using a linear function of variance is precise only in the continuous-time case when return moments higher than the variance vanish. Consequently, for some investor circumstances, higher return moments may need managing in any return environment.

2.2 Discretionary Wealth Model and Applications

We reframe growth optimality to apply not to total wealth, nor even to investments, but to a kind of implied balance sheet surplus we will term *discretionary wealth*. In doing so, we answer earlier objections to growth-optimal normative models. For some readers, the change in frame of reference may initially be disorienting. In the practice of astronomy, the relative motions of the planets could be far more simply explained after Copernicus and Kepler reframed

---

[5] This seems to be a misunderstanding based on some ambiguous remarks in footnote 5 of Fama (1968): *"Tobin claims (and properly so) that the mean-standard deviation framework is appropriate whenever distributions of returns on all assets and portfolios are of the same type and can be fully described by two parameters."* This is in turn seems to be based on a clearer and less sweeping observation by Tobin (1958): *"Whatever a two parameter family is assumed —uniform, normal or some other — the whole probability distribution is determined as soon as the mean and standard deviation are specified. Hence the investor's choice among probability distributions can be analyzed by μ_R-σ_R indifference curves."*

them from epicycles around circles around the earth to ellipses around the sun. In similar, though perhaps less cosmic, fashion we reframe risk aversion. The convention has been to plot "utility" against wealth, representing different degrees of risk aversion with different degrees of curvature. Instead, if we were to express our normative approach in utility terms, we would employ only a single logarithmic curve. Since the full range of curvatures is monotonically distributed along the logarithmic curve, we can represent any degree of risk aversion through changing the curve's scaling. For example, if the value of investments is three times that of discretionary wealth, then a 10% loss in the investment portfolio is a 30% loss in discretionary wealth, placing the interval of change in a region of greater curvature, giving rise to an amplified aversion to risk. Consequently, our model objective for discretionary wealth is to maximize the expected $\ln(1+Lr)$ each period, where L is the ratio of investable assets to discretionary wealth.

Who determines discretionary wealth? The investor determines it directly, and if one wants to think in terms of utility, it is discretionary wealth plus its expected growth criterion, or the implied goal of long term best median wealth without intermediate shortfall, that determines utility, rather than the other way round.

We define discretionary wealth in terms of an accounting balance sheet, as in Table 1. On the left side are our investment assets, plus the time-discounted value of foreseen financial contributions to the portfolio. This present value is an implied asset. On the right side are our current debts, plus the present values of our foreseen financial commitments that must be satisfied by withdrawals.   These latter make up an implied liability. The residual or surplus on the right side we term discretionary wealth. In practice, there will be a fuzzy boundary between implied liabilities that are truly committed and what remains as surplus. Here in Section 2, the example amounts are sharply defined. In Section 3, we will allow balance sheet quantities to have probability distributions. We also assume that we have captured all the foreseeable cash flow, and that future cash flow surprises will be independent and bounded, in parallel to our assumptions for investment returns. This assumption preserves the approximately recursive nature of the investment policy to be derived.[6]

<INSERT TABLE 1 ABOUT HERE>

In the context of the iterative process for coordinating consumption and investment policy noted earlier, we further assume that forecast spending and saving have already been adjusted so that discretionary wealth is positive, and we need only consider how to make it grow. An emphasis on growing something on the right-hand side of the balance sheet may require further mental accommodation; discretionary wealth in accounting terms is not an asset but rather more akin to book equity. It may be distinctly suboptimal to allocate an amount to risky investments precisely equal to our discretionary wealth.

2.2.1 *Application to quantifying the impact of higher moments*

---

[6] There is a second-order effect spoiling complete independence of portfolio level returns that prevents perfect recursion. Future appropriate risk aversion is contingent on current changes in discretionary wealth, thereby affecting future return volatility at the portfolio level. While valuable in reducing the probability of shortfalls, this effect leaves open the theoretical possibility of still better multi-period strategies.

In Table 1, the ratio of the value of the investment portfolio to that of the surplus or discretionary wealth is 2.5. We refer to this as *implied leverage*, or *L*, though no borrowing is involved, only scaling of return. That is, in this example if we have a 10% loss in the investment portfolio, it results in a 25% loss in our discretionary wealth. Both the mean and the standard deviation of arithmetic return are multiplied by the implied leverage *L* when we remap them from the investment portfolio to discretionary wealth. This means that variance is multiplied by *L* squared; the third moment, involving skewness, is multiplied by *L* cubed; the fourth moment, involving kurtosis, is multiplied by *L* raised to the fourth power. The Taylor series for mean log return has been mapped from the investment portfolio to discretionary wealth; it is shown as equation (3).

$$\text{Expected } \ln(1+Lr) \cong \ln(1+LE) - \frac{L^2V}{2(1+LE)^2} + \frac{SL^3V^{\frac{3}{2}}}{3(1+LE)^3} - \frac{KL^4V^2}{4(1+LE)^4} + \dots \tag{3}$$

Equation (3) makes very clear the functional relationship of expected compound growth as it depends on leverage. Begin with variance *V* considerably smaller than expected return *E*. As implied leverage *L* is increased, the initial impact is to improve compound return, but with further leverage increases this relationship eventually peaks and turns downward. With still further increases in leverage, expected log return for discretionary wealth becomes negative and ensures eventual shortfall.

It should be clear that even if the investor has correctly specified expected log return on discretionary wealth to the best of his or her knowledge, it is always possible, though unlikely, for actual returns to produce negative discretionary wealth. In the context of recent market disruptions, perhaps the more relevant question for many is how one avoids a situation where a market crash makes it impossible to finance obligations. The discretionary wealth approach itself provides a mild form of dynamic hedging that reduces this probability. The fuller Bayesian model in Section 3 reduces it further because a wider spectrum of possible disturbances in a posterior distribution for mean and covariances is taken into account. For example, if the posterior distribution for the mean is a Student's *t* distribution and that for covariances is an inverse Wishart distribution, the resulting prediction distribution for returns will have fatter tails than the data that informed the posterior distributions. In addition, one can introduce reserves for unknown contingencies as an additional implied liability. This allows us to get the benefits of dynamic adjustment of risk aversion as we reach lower levels of discretionary wealth. The conventional practice of dedicating risk-free assets sufficient to cover all liabilities, as in some pension asset-liability management and insurance accounts, is demonstrably suboptimal unless transaction costs are very high. However, if discretionary wealth nevertheless becomes negative, one can adjust only by increasing saving and decreasing spending plans. Ideally, as in Section 2.2.3, this will have been done earlier as implied leverage increases to concerning levels as part of solving the more global financing problem, but again there is no foolproof answer.

Others have investigated portfolio optimization using the first three or four return moments of a Taylor series approximation.[7] However, our focus on expected growth of log

---

[7] See Harvey *et al* (2002), for example.

return on discretionary wealth goes further. It better identifies the investor contexts where this might be worthwhile and supplies through equation (3) an appropriate risk aversion for each return moment. That is, dividing through by L to put ourselves in the Markowitz context, the appropriate risk aversion for the second return moment is approximately $L/2$, the appropriate risk aversion for the third moment is approximately $L^2/3$, and the appropriate risk aversion for the fourth, or kurtosis, moment is $L^3/4$.

Equation (3) also highlights the combination of high leverage, high variance, skewness, and kurtosis as joint determinants of the relative strength of higher moments. This expands the usual research interest in return skewness and kurtosis to a broader list of interacting factors, making clear the relevance of higher moments to households with marginally adequate investment assets to support their future spending commitments. Return higher moment terms can become material relative to the variance term in a rather nonlinear and explosive fashion as discretionary wealth is shrunk, even for a portfolio that appears conventional. Though the investor can frequently limit the relative impact of higher return moments by shortening the length of periods between investment decisions, this ability may not always be practical, especially for the household investor.

Finally, if one can estimate skewness and kurtosis, equation (3) provides an excellent way to measure the impact of downside risk not captured by variance, reducing the need for measurement of semi-variances and assorted *ad hoc* ratios.

2.2.2 *Application to Markowitz optimization*

The exploration of appropriate risk aversion noted above can be translated directly to the Markowitz framework with minor approximation. Suppose that LE and L^2V are small. Then an approximation of equation (3) implies that:

$$\text{Expected } \ln(1+Lr) \sim LE - L^2V/2 \tag{4}$$

In cases where L is fixed, whatever investment portfolio optimizes the value on the right-hand side of equation (4) will also maximize the value on the right-hand side of equation (5).

$$(1/L) \, [\text{Expected } \ln(1+Lr)] \sim E - LV/2 \tag{5}$$

This brings us back to the objective function of Markowitz mean-variance optimization, with $L/2$ providing the risk-aversion coefficient. In that context, we have specified optimal risk-aversion, or where the investor should be on Markowitz's efficient frontier of best return for a given degree of risk. This result, providing Markowitz's missing normative risk aversion, is of great practical significance.

2.2.3 *Application to planning consumption*

Suppose in contrast that the investment characteristics of the portfolio are fixed. What can we say about desirable changes in the implied leverage? Solving for the optimal L in equation (4), we arrive at $L^* = E/V$. This should be sought through adjusting consumption plans.

However, if the allocations to investment assets are not fixed, we can determine only the best products of implied leverage and asset weights. We cannot say without further information

in the individual case whether it is *L*, or instead *E/V*, that should be altered. However, for each combination, we can estimate the projected mean growth rate obtained by substitution into equation (4) as $(E^2/V)/2$. The optimal growth rate thus estimated, combined with the implied coordinating consumption plan, should be useful information to the investor in choosing among the available scenarios.

2.2.4 *Application to taking taxation into account*

Because the political process produces complicated tax rules, a detailed treatment of the application of the discretionary wealth paradigm to taxable investors is too extensive to attempt in this paper. However, broad guidelines can be noted.

First, portfolio optimization should always be done in after-tax terms. It should be noted that this is not yet generally done in practice. Equation (3) makes clear that different effective tax rates fundamentally change investment characteristics. Assuming gains and losses can be offset, taxation of returns ─ whether interest, dividends or capital gains ─ affects both mean and standard deviation of returns by multiplying them by $(1 - T)$, where *T* is the effective tax rate. Then taxes affect variance by $(1 - T)$ squared, and affect the higher return moments by greater powers of $(1 - T)$. The difference influences appropriate allocation. High-tax-bracket investors, other things equal, should appear less risk averse than low-tax-bracket or tax-exempt investors, and especially less averse to higher moment risk.

Second, because some taxes, at least in the US, are levied at a point in time much later than the event that triggers the liability, there is additional capital available for compounding until tax payment. This can make the effective tax rate for long-held assets that increase in value considerably lower than the nominal long-term capital gains tax rate. Also, because of the possibility for what is known as tax loss harvesting, the tax-timing option values created for volatile securities can further increase the additional capital available for compounding. One way to approximate this effect for such assets is to further reduce the effective tax rate applied to equation (3).

Third, taxes already owed may be assigned present values as implied liabilities. For example, one may partition capital gain taxes on unrealized gains into the part corresponding to current prices, represented as an implied liability, and that which may be incurred from future gains, represented through adjustments to return parameters in equation (3). Similarly, an unrealized loss may create an implied tax-benefit asset. For those wealthy enough to incur them and who care about the after-tax size of their estate, in principle estate taxes may be treated similarly by partitioning their representation between implied liabilities and impact on future return parameters.

2.2.5 *Application to dynamic hedging and bubbles*

The simplest form of dynamic hedging is CPPI as described in Black and Perold (1992). Too often the CPPI "multiplier", which here plays the role of leverage as it relates to the "cushion" above a "floor," has been set by commercial vendors rather high so as to produce a more appealing positively skewed return. But if the multiplier is set too far above *E/V*, and if the first few periods are unlucky so that constraints on risky positions do not soon restrict the

16

effective multiplier to a value closer to optimum, the expected log return on the cushion may start negative and remain negative, and consequently the investor may become stuck at the floor.

Recursively maximizing the expected log return on discretionary wealth can result in a mild form of dynamic hedging. It should be recognized that so long as the causes of changes in discretionary wealth are common across investors, and investors with low leverage and with high leverage do not act to offset one another, this household optimizing behavior will lead to increased market volatility. This is a "tragedy of the commons" problem relevant to those concerned with regulating markets.[8] While maximizing expected log return on fluctuating discretionary wealth can be very beneficial for risk control for the individual household, when undertaken collectively in the same direction, it will tend to reinforce bubbles and panics.

2.2.6 *Application to life-cycle investing*

If we think of risk aversion as rooted in personality characteristics, the profile questionnaires often used by financial advisors in establishing customer accounts do not encourage adjustments based on life contingencies. In contrast, the discretionary wealth approach recognizes differences both in current implied leverage, including factors that affect the present value of future retirement spending. This not only helps customize an investment strategy to investors, but it provides for path-dependent adjustments in asset allocation throughout a life-cycle. The appropriate asset-allocation path from youth through old age may be quite different for investors with different wealth endowments and participation in different investment environments, calling into question currently popular life-cycle mutual funds whose asset allocation trajectory or "glide path" is based on years to a fixed date and that are marketed to investors depending on age of retirement.[9]

In the discretionary wealth approach, the typical changes in implied assets and implied liability as the investor approaches retirement and then approaches the end of life provide a natural evolution in asset allocation. However, the factors of initial wealth endowment, investment results and life events such as marriage, children, and health will also be reflected, so that investments can be far better customized to the household's needs.

As noted earlier, at some point close to the end of life, the shorter number of periods ahead imply that the maximization of mean log return on discretionary wealth may be noticeably improved upon as a means of maximizing the median ending result. One can imagine dynamic programming solutions. However, the discretionary wealth model can be applied in truncated two-moment form at that point, so that it will always be as good as Markowitz mean-variance optimization enhanced by an appropriate risk aversion.

---

[8] Hardin's "The Tragedy of the Commons" describes a dilemma in which multiple individuals acting independently in their own self-interest can eventually destroy a shared resource despite the fact that it is not in anyone's long-term interest for this to occur (Hardin, 1968). This highly influential writing has been applied to a wide range of fields that deal with resource allocation problems such as ecology, overfishing, and economics. In economics it is used to describe the failure of incentives in the private sector to provide satisfactory maintenance of public resources.

[9] We apologize to the reader who would like to see more detail on this important practical topic, and hope in the future to provide examples space does not permit here.

2.2.7 *Application to performance measures*

Performance measures used by institutional investors to monitor investment managers are typically focused on detecting the skill (or its lack) of professional investment managers, as in Sharpe (1966, 1994). Equations (3), (4), and (5) make clear that comparisons of good and bad performance can be better customized to the risk aversion needs of the particular household. Such a customized performance indicator, aside from any adjustment for sensitivity to higher return moments, is a comparison of $E - LV/2$ for the investor against $E' - LV'/2$ for the benchmark. The implied leverage L, customized to the investor, is held the same for both the benchmark and the portfolio. This is clearly superior to the "information ratio", which discriminates against risk-reducing strategies. In many cases it will also be more useful than the Sharpe ratio. Only when implied leverage itself is near optimal, that is $L \sim L^* \sim E/V$, will a ratio of E to $V^{1/2}$, such as implicit in the Sharpe ratio (ignoring the usual netting of the risk-free rate), judge portfolio performance in a manner nearly optimal for household use.[10]

If higher moments are important, a more comprehensive performance measure can be easily derived as a comparison of average logarithmic leveraged returns, using the investor's implied leverage. That is, we compare average $\ln(1 + Lr)$ to average $\ln(1 + Lr')$.

2.3 Hypotheses for Descriptive Research

Consider that the maximization of long-term median wealth implies that investors following a policy of maximizing expected log return on discretionary wealth will tend in aggregate to accumulate disproportionate wealth. It is a short step to hypothesize that the overall market structure is influenced by the preferences implicit in equation (3). This would help explain phenomena that might appear mysterious within a narrower model. These come in three categories – pricing phenomena explained by normative risk aversion to variance, pricing phenomena explained by aversion to higher return moments, and clientele effects explained by differences in implied leverage.

2.3.1 *Application to variance risk aversion*

It is not entirely clear whether it is a bond risk premium or a stock risk premium that constitutes a puzzle for market models of pricing, (Mehra, 2007). However, according to the discretionary wealth model, the long-term risk premium associated with the additional return of stocks and bonds over that of nearly riskless investments ought to be influenced by a weighted average of implied leverage across market participants, in turn associated with balance sheets augmented by implied assets and liabilities. Empirical estimation of investor augmented balance sheets (along with other model factors such as negative skewness in returns) might help explicate what have sometimes been regarded as puzzles at the apparent size of such premia.

Also, over time, aggregate adjusting of allocations to reflect changing risk aversion resulting from changes in household balance sheets would tend to result in the kind of price

---

[10] To see this, substitute L^* into equation (4). If the risk-free rate were zero, half the squared Sharpe ratio would give the optimum growth rate of discretionary wealth.

volatility viewed as anomalous by Shiller (2000). This phenomenon should contribute to widespread asset price bubble formation, with or without speculation as to future prices.

2.3.2 *Application to aversion to higher moments*

Many additional stylized facts often viewed as anomalies ought to follow from equation (3)'s inclusion of higher-order return moments involving skewness and kurtosis. As earlier noted, equation (3) reduces to a relationship involving only mean and variance when independent returns are combined with continuous decision-making. However, when realistic trading and decision-making costs are added, lengthening time periods between reallocations, as well as implied leverages in many cases considerably greater than unity, make higher-order return moments increasingly important. Securities and trading activity that provide additional positive skew or reduce kurtosis become more valuable. Here are some of the more salient implications if this value is priced.

Active investors following a "value-oriented" policy and active investors following a "momentum" policy take opposite sides of a trade in high-order return moments. A momentum policy buys positive skewness through dynamic hedging and a value policy sells positive skewness, as described by Leland (1999). Consequently, value investors should earn higher returns as compensation. It is unclear whether this is in addition to, or subsumed by, a premium equivalent to the cost of options offering similar protection, but in any case higher expected value-oriented returns should exist, as reported by Fama and French (1992).

Small capitalization stocks may be small because of low market to book ratios and consequently earn a value premium. But in addition, most very small capitalization stocks tend to be less liquid. Higher transaction costs effectively lengthen the time between investment decisions and increase the relative importance of higher return moments. This increases the relative importance of return kurtosis and provides a basis for an illiquidity premium consistent with the small-stock premium documented by Fama and French (1992).

Investor lock-ups, such as those for venture capital partnerships with 10-year restrictions on withdrawals, more generally may be supposed to earn an illiquidity premium reflecting exposure to return kurtosis.

There is already considerable evidence for pricing of skewness in portfolios, as in Harvey and Siddique (2000). But the most persuasive evidence may be in option pricing. By definition, out-of-the-money options relate to the underlying security's higher return moments more than do at-the-money options. Out-of-the-money puts and calls provide purer protection against negative skew and should therefore be valued more highly than implied by the Black-Scholes continuous trading arbitrage pricing model. This standard model is based on an assumption of option value as protection solely against variance. Its application to a spectrum of option prices should give rise to the so-called volatility "smile."

Similarly, the simultaneous sale of an out-of-the money call and purchase of an out-of-the-money put protects against kurtosis in the return of the underlying security. Since this is valuable, the put should be priced at a higher implied volatility than the call, giving rise to the volatility "smirk" when option pricing is interpreted as implied variance within the Black-

Scholes model. Note that this result may be non-intuitive: the probability distribution symmetry implicit in kurtosis gives rise not to the symmetric "smile" but rather to the lopsided "smirk."

2.3.3 *Application to clienteles*

There are sound reasons for investor clientele niches unexplained within the standard Capital Asset Pricing Model (CAPM). We refer to niches above and beyond those engendered by the existence of tax-exempt municipal bonds or differences in information availability. Clientele niches can also arise from differences in appropriate response to higher-order return moments. Consider, in the context of equation (3), the effective shrinkage of return standard deviation both from higher tax rates and separately from lower household balance sheet implied leverage. Because third and fourth moments involve increasing powers of standard deviation, investors with higher after-tax implied leverage should be less willing to hold securities or take actions that involve more negative skew and higher kurtosis.

Because most investors cannot make even approximately continuous decisions, here we have a definitive break from the foundations of market models based on continuous finance, or on any model in which only the mean and variance terms are relevant (e.g., CAPM, Sharpe, 1964). If surviving and dominant investors tend to act as suggested, then a two-fund separation theorem, as in Tobin (1958), will be a poor guide to good investment policy. This has broad implications for the kinds of investment products investors with different implied leverages should hold. For example, a retiree requiring income to support a minimum lifestyle, consequently having high implied leverage L, should not write call options, adding negative skew to his or her portfolio. It is the wealthy investor with low implied leverage who can best afford to write options resulting in negative skew, or is qualified to take on a 10-year lockup (with kurtosis effects) for a venture capital fund. Neither should invest their risk-bearing assets exclusively in a conventional one-size-fits all index fund. It will neither fully protect the marginal retiree against higher return moments nor extract a full risk premium from them for the wealthy investor with ample discretionary wealth.

2.4 Simple Example Using Point Estimates

A 55 year-old woman with a cash inheritance, after-tax, of $10 million desires a $300,000 income after-tax for her remaining lifetime, in constant dollars after inflation.[11] An after-tax, after-inflation interest rate is estimated at 2.5%. Her actuarial life expectancy is 27.7 years. A present value calculation of this implied liability as an ordinary annuity is about $5.95 million. Her discretionary wealth calculated on the basis of mean lifetime would be $4.05 million, and her implied leverage would be 2.47. A more relevant point estimate, however, is that for the mean implied leverage, not the implied leverage calculated from mean lifetime. Taking into account the probabilities for each possible lifetime from the actuarial tables, and using those to calculate a probability distribution for implied leverage, the leverage distribution is skewed with a long tail toward higher leverage as discretionary wealth is reduced. The mean implied leverage is a somewhat larger ratio of 2.72.

---

[11] This case is adapted from Wilcox (2008).

We consider a case where the investment decision is not affected by transaction costs, and the decision is what weight to give to four assets, a cash equivalent, a bond, and two stocks. No short positions are allowed. The expected after-tax returns are, respectively, 2%, 3%, 6% and 6%. The after-tax standard deviations are 1%, 5%, 20% and 20%. There is no covariance of cash and bond returns with anything else, and the two stock returns have a 0.9 correlation. Given a simplified two-moment objective of maximizing expected portfolio return less half the implied leverage of 2.72 times the portfolio variance, the best asset weights are about 0%, 66.7%, 16.7%, and 16.7%. We have customized the degree of conservatism to the investor's needs. In the next section we will see how to improve this fit and how to accommodate more difficult cases. The example is extended to take into account imprecision in these decision-input parameters through the use of Bayesian logic in Section 3.3.

3. BAYESIAN INVESTING

We begin with the universality of imprecise knowledge and adapt our investment policy to that assumption. Consequently we embrace Bayesian logic.

The Bayesian revolution in probabilistic reasoning seems to be gradually diffusing into financial research.[12] This is not limited to methods for more robust derivation of point estimate return parameters for use in conventional Markowitz mean-variance optimization. With the retention of full probability distributions for intermediate calculations, the subsequent asset allocation process itself can be upgraded.[13] One benefit is greater robustness to errors in sample estimates. A second benefit is greater appropriateness of choices to available information through taking into account nonlinear transformations of probability distributions. Further, once Bayesian methods are mastered, it is easy to include uncertainty as to the appropriate risk-aversion parameter, for example that resulting from longevity risk, as a factor influencing the optimal portfolio allocation.

3.1 Better Point Estimates for Markowitz Inputs

We can improve conventional Markowitz mean-variance optimization by using Bayesian methods even if they are employed merely to derive better point estimates for input parameters. The well-known shrinkage models for expected return and covariance are, in spirit, highly specialized Bayesian models adapted to derive better point estimates for input parameters.[14]

---

[12] For a review of basic Bayesian concepts and methodologies, see Gelman *et al* (2004) and Rachev *et al* (2008). It and similar texts describe Bayes' Law, conjugate prior and posterior distributions for representing knowledge about likelihood distributions, and Markov chain Monte Carlo methods such as Gibbs sampling for more efficiently producing posterior distributions in complex situations. For texts on MCMC, see Roberts and Casella (2005) or Liu (2008). This is still a fast-moving field in terms of techniques for efficient handling of larger and more complex problems; an Internet search using terms like "reversible jump MCMC", Green (1995), "particle filters", Liu (2008) and "quasi-Monte Carlo", Tribble and Owen (2008), will assist the reader in keeping up with recent progress.

[13] The literature here has become too large to do justice to all the relevant research. Meucci (2005) and Fabozzi *et al* (2007) review a wide span of applications. Recent contributions incorporating a Bayesian decision include Greyserman *et al* (2006), Kan and Zhou (2007), and Harvey *et al* (2008).

[14] This thread includes, for example, Jorion (1996), Black and Litterman (1992), and Ledoit and Wolf (2004).

Here we instead briefly review several more general Bayesian techniques for better point estimates of return distribution parameters, each having the useful, but too little used, by-product of making available posterior probability distributions. Since these techniques are reported elsewhere in the statistical literature, we will save our examples for the more subtle improvements in subsequent decision-making to be described after introducing our second model.

3.1.1 *Application to risk estimation for portfolio construction*

Prior to the development of more capable Markov chain Monte Carlo (MCMC) methods in the 1980s, Bayesian estimation was typically done within *conjugate* probability distributions. That is, we represent our knowledge with prior and posterior distributions, where both are assumed to be from the same function family, differing only in parameters that increment by specific formulae as more data are added. Commonly used likelihood distributions are each associated with their own conjugate prior-posterior distribution family. The Bayesian distinguishes between, first, the data-generating process as likelihood distribution and, second, our knowledge of it as prior and posterior distributions. This distinction has important implications for portfolio construction.

For example, suppose that a return-generating process is believed to be normal, but with unknown mean and unknown variance. The relevant data are limited. Regrettably, the normal distribution as prior and posterior is not conjugate to a normal likelihood distribution. A different distribution — a normal mean conditional on a scaled gamma distribution for the inverse of the variance — is conjugate.[15] It leads to a Student's *t* distribution for the marginal distribution of the mean — which can produce fat tails. Also, the marginal distribution for variance will be skewed, with a fat tail extending to higher values. That is, no matter what variance we have observed, there is always a possibility that these observations were drawn from a likelihood process with much larger variance, whereas it is impossible for the likelihood variance to go as low as zero if we have observed variation. Ignoring these two facts by focusing just on observed sample variance, along with the usual omission of the variance of the mean, leads to both variance and kurtosis underestimation. In contrast, following Bayesian logic by first generating a prediction distribution for returns based on random draws from posterior distributions for mean and variance, and only then measuring the parameters of that distribution, avoids these sources of bias.

3.1.2 *Application to hierarchical models for return estimation*

Consider a household investor attempting to discern investment skill among many investment managers or advisors with similar described approaches but different adjusted performance records. Absent more detailed knowledge, we can improve our assessment of each manager's skill by modeling this situation with a *hierarchical* likelihood. For example, we might assume the likelihood of each individual manager's mean return is normally distributed around a global mean with a global variance. For each of the managers, that manager's observed

---

[15] See Gelman *et al* (1995).

return is normally distributed around that manager's mean return, according to his or her individual return variance.

The Bayesian estimation of the implied parameters shrinks the otherwise widely dispersed estimates of each manager's mean return toward the group average. The greater the number of potential managers, other things equal, the greater will be the shrinkage. The application of this principle, which underlies the appreciation of survival bias, would seem to us to have societal benefits.

Consider further the problem of the quantitatively-oriented active investor seeking to develop return forecasts for stocks in many industry sectors. For example, one might seek return forecasts contingent on advantages in forecasting the price of oil. We are given a time-series of past changes in forecasts of oil prices, together with monthly return data on a thousand common stocks for a 10-year history. We could pool these data and estimate a single ordinary least squares regression relationship, but we suspect that the slope of the relationship may differ by industry sector. We proceed to estimate a hierarchical regression model. That is, individual true sector regression coefficients are drawn from a global distribution, and the observed coefficients for each sector are drawn from the individual sector distributions of observed regression coefficients around their true individual sector regression coefficients. MCMC methods are particularly useful for such estimations.[16]

3.1.3 *Application to forecast model averaging*

Consider the following likelihood linking two separate models, in this case incorporating either X_1 or X_2, along with uncorrelated noise terms ε, as predictors of Y.

$$Y1_t \sim a_1 + b_1 X_1 + \varepsilon_1 \tag{6}$$

$$Y2_t \sim a_2 + b_2 X_2 + \varepsilon_2 \tag{7}$$

$$Y_t \sim \text{Binomial}(\theta, Y1_t, Y2_t) \tag{8}$$

Equations (6) and (7) are alternative linear relationships. Equation (8) expresses the idea that for any given observation, there is a probability θ that $Y1_t$ rather than $Y2_t$ is the true underlying model. Given posterior distributions for θ as well as for both the individual model components, we can construct a prediction distribution for Y.

The Bayesian statistics literature notes the increase in prediction accuracy possible through this kind of model averaging rather than through a focus on the single best model.[17] For investment purposes, there is particular benefit through incorporating additional information in cases where correlated predictors pose significant statistical problems for classical econometric methods.

---

[16] An easy introductory tool is WinBUGS, Lunn *et al* (2000).

[17] See George and McCulloch (1993) and Carlin and Chib (1995). An alternative approach that is popular is reversible jump MCMC (also called the dimensional MCMC), see Green (1995).

To provide a concrete example, suppose one is forecasting currency returns based on interest rate differentials across countries (ignoring autocorrelation issues). The cross-country differences in yields in one-month maturities and one-year maturities each appear to be a significant positive predictor of subsequent weekly currency returns. However, these two predictors are themselves very positively correlated (0.98). If we put them both in a conventional ordinary least squares regression, we will arrive at a strong positive coefficient for one predictor and a strong negative coefficient for the other. However, such models do not generalize well to future data. Model averaging in this case produces a prediction distribution reflecting positive coefficients for both predictors, greatly reducing sensitivity to small-sample error.

Model averaging is potentially of very wide applicability. Unlike hierarchical models, which capture nested relationships, model averaging is a general approach for combining non-nested relationships. This concept extends not just to alternative predictive variables, but to entire systems of modeling, such as one faces in combining industry sector, country and style factors in forecasting international stock returns.

3.2 <u>A Fully Bayesian Model</u>

We can extend Bayesian logic beyond initial input estimation to subsequent *decisions*. Our second model looks the same as before: maximize each period expected $\ln(1 + Lr)$, the expected logarithmic return on discretionary wealth. The novelty of this model, however, is that a joint prediction distribution for the components of $\ln(1 + Lr)$ is created before making a decision as to best portfolio allocation to investments. The hallmark of a fully Bayesian decision is the preservation of all relevant information until the final ranking of decision alternatives.

Our approach should not be confused with a "resampling" method that deals with uncertainty by building a portfolio based on an average of security weights derived through repeated conventional mean-variance optimizations given perturbed input parameter point-estimates.[18]

3.2.1 *Why preserve input and intermediate probability distributions?*

Point estimates are so much easier to use, and consequently used so frequently in everyday financial calculations, that we may forget they are mere abbreviations of our knowledge. Our discussion in Section 3.1 of fully Bayesian decision-making implies that it is a mapping from input probability distributions and decision alternatives to output probability distributions, the latter further mapped by a functional into single-dimension preferences. In contrast, conventional practice with Markowitz mean-variance optimization collapses information to point estimates at the beginning. That is, a functional transformation of information from higher dimension to lower dimension takes place before calculations of the

---

[18] Michaud (2001) has shown that averaging classically optimized portfolios indicated by perturbed input point estimates can improve results. This appeared to Markowitz and Usmen (2003) to be better than a Bayesian approach, though with a rejoinder by Harvey *et al* (2008). Although a fully Bayesian decision model can involve considerably more computation than the resampling approach, it avoids being misled by problem features involving nonlinear calculations.

optimal security weights. Whether the discarded information is relevant or irrelevant to this evaluation is greatly influenced by the degree of linearity of the mappings in the process.

Consider the following examples of potential mappings involving nonlinearities. The mean of the product of two normal distributions is not the product of the component means. The mean of the distribution for $1/X$ is not the reciprocal of the mean for X. The means of the distributions for each of the elements of a covariance matrix's inverse cannot be determined from the means of the distributions for each element of the original matrix. The mean of a constrained variable's distribution cannot be obtained by constraining the mean of its original distribution. In all these non-affine (nonlinear) cases, an inference as to the mean of the function's results is biased by attempting to infer it by applying the same function to the mean of the input distribution.

The presence of mapping nonlinearities before the final preference functional calls into question any decision based on point estimates of inputs with dispersed probability distributions.

3.2.2 *More comprehensive picture of optimization nonlinearities*

Suppose the returns of two financial assets differ in their means and variances by only infinitesimal amounts, and that the correlation of their returns is 0.99. If we optimize using full probability distributions with realistic predictive dispersion, their appropriate weights will be close to equal. In contrast, if we use conventional point estimates to make the calculation, we approach a singular matrix. We will then choose to take a very large long position in one security and a very large short position in the other. This is an invalid conclusion to a simple situation.

Two sequences of mappings are involved in comparing conventional Markowitz optimization to Bayesian optimization of mean log return on discretionary wealth; they are each highly nonlinear.

Consider first the conventional use of Markowitz mean-variance optimization. The mapping of investor factors required to arrive at a fixed risk-aversion parameter is unknown and we will put it aside. The remaining portions of the mapping sequence runs from input distributions for return characteristics to input point estimates, and thence to a proposed optimal set of weights, omitting consideration of a probability distribution for the objective function.

In contrast, the Bayesian mapping sequence runs step by step from funding input distributions through present value and leverage ratio distributions and step by step from posterior distributions for return characteristics to predictive distributions for return, often with significant higher moments. We derive a probability distribution of the objective function, including all constraints, as a function of weights and thence arrive at an optimal set of weights.

Because of nonlinear mappings within both sequences, the results will usually diverge, sometimes very little, but sometimes disastrously. Let us consider in more detail the original problem of Markowitz as expanded to show its relationship to expected log return. It maximizes:

Objective = $\max[Lw'E - (L^2/2)w'Vw]$, subject to constraints. (9)

25

Equation (9) can be regarded either as a relationship of point estimates or of probability distributions. Conventional practice takes it as relating point estimates. In equation (9), w (the vector of security weights), is subject to at least a budget constraint, and possibly to other constraints, w' is its transpose, E is the vector of point estimate mean security returns, and V is the matrix of point estimates of return covariances. L is a risk-aversion parameter, which, since it is conventionally viewed as a point estimate, can be divided through to produce the familiar form of Markowitz's objective function.

Suppose the only constraint was the budget constraint; the sum of the weights, positive or negative, must be unity. We can solve for optimal weights using a Lagrangean multiplier for this constraint. Augmenting V to V'' and E to E'' to take into account the multiplier, we produce the optimal weights to the point estimate version of equation (9), augmented by a value for the multiplier, to form w'' as in equation (10).

$$w'' = V''^{-1}(E''/L) \tag{10}$$

We see that the determination of weights in equation (10), whether explicitly accomplished as here or implicitly arrived at through the use of penalty functions or some other method, involves the inverse of the augmented covariance matrix, a highly nonlinear operation. This is multiplied, another nonlinear operation, by mean returns, and divided, another nonlinear operation, by the implied leverage determining appropriate risk aversion.

Consider now the Bayesian sequence of mappings of probability distributions from inputs to the probability distribution of the objective function as a function of security weights, beginning with equation (9) as relating probability distributions rather than point estimates. The initial calculation of the probability distribution for L involves nonlinearity in both (1) present values to determine assets and liabilities and (2) a leverage ratio whose denominator discretionary wealth is a difference. The squared value of the distribution for L induces another important nonlinearity, especially if the distribution for L has a long upward tail, as would be the case for a moderate income retiree who might live longer than expected. Since neither E nor V is known in reality with much precision, and since the appropriate posterior distribution for V is skewed with a long upper tail, the effective scope for departures from the point-estimate-based method is increased further by multiplication to produce distributions for LE and L^2V, respectively.

Comparing the two mapping sequences, a conventional point-estimate-based sequence of nonlinear calculations cannot be expected to closely track the result of compounding probability distributions through the intervening nonlinear mappings. Those who use such conventional optimizations must employ a variety of *ad hoc* strategies to produce acceptable practical results.

Adding position constraints, such as the requirement that no securities take negative weights, and that no security can take more than a modest maximum weight, can reduce the biases introduced by the point estimate approximation. Markowitz (1983) did something similar when he advocated allowing only constrained short positions. However, at best, constraints reduce errors in suboptimal fashion. Additional strategies, such as (1) reducing the number of securities considered or (2) assuming a simpler factor structure for V (as in Jacobs *et al*, 2005), or (3) Bayesian shrinkage of the input V toward a simpler structure (as in Wolf and Ledoit, 2004),

can all reduce the scope for distortion by reducing the nonlinearities produced by matrix inversion. However, though these strategies reduce the symptoms, they do not truly cure the disease of relying on point estimates for a problem that includes non-affine mappings of probability distributions. In any case, the disease reappears as soon as we include uncertain risk aversion in our problem.

3.2.3 *Application to integrating appropriate risk aversion uncertainty*

Since Markowitz's 1952 contribution, financial research has paid a great deal of attention to risk regarding investment returns. In contrast, far less attention has been given to the inevitable, and often quite large, risk concerning the household investor's future sources and uses of investment funds. The present values of these future cash flows are important determinants of current discretionary wealth, and in turn determine the appropriate level of single-period risk aversion. Consequently, appropriate risk aversion should not be considered as a point estimate, but as a probability distribution. If this is not done, optimization will not know enough to avoid unfortunate *Lr* combinations: for example, longer-than expected life, resulting in high implied leverage, combined with worse than expected returns. When we consider appropriate risk aversion as a probability distribution, we further discover that the shortcut Markowitz objective function obtained by dividing equation (4) by *L*, the implicit leverage, is often inadequate for household investment. Whenever, for example, the mean of *L*-squared is not very close to the square of the mean of *L*, the Markowitz objective will tend to produce additional distortions away from our objective. This effect tends to be exaggerated for those whose discretionary wealth is slim. Fully Bayesian decision-making applied to the discretionary wealth approach gives us a conceptually simple way of bringing these additional considerations into the household investment decision.

Uncertainties with regard to future cash contributions and withdrawals include not only longevity risk but many other sources and uses of investment funds. Additionally, future cash flows must be time discounted to create implied balance sheet items. The associate risks can of course be addressed by adding hedges to the portfolio problem, as when a pension fund considers Treasury inflation protection securities (TIPS) rather than Treasury bills. However, this consideration of hedges is more accurately done in the context of a more complete representation of the problem. Uncertainty as to the time discount rates to be applied to both real plans, such as maintaining a standard of living in retirement, and nominal plans, such as paying a capital gains tax on today's value of stock holdings, can be a major cause of uncertainty in implied assets and liabilities.

All this complexity can be summarized in a single distribution for discretionary wealth. Even so, the subsequent mapping of uncertainties in future funding flows into discretionary wealth and thence into the distribution for implied leverage *L* is highly nonlinear, making it especially important that we address it. Imprecision in appropriate risk aversion does not prevent us from portfolio optimization. However it is a further reason for households facing uncertain needs not to place reliance on conventional mean-variance optimized portfolios. Anecdotally, many financial advisors do use these models.

3.2.4 *Application to fully Bayesian optimized portfolios*

Conceptually, the fully Bayesian portfolio construction decision process is simple, though it may involve a large number of calculations. We need to begin with likelihood process for a distribution of returns.[19] Suppose we have already available joint posterior probability distributions for return mean vector *E* and covariance matrix *V*.[20] We then randomly draw from it a large sample of *E,V* pairs. These pairs are then substituted into our likelihood model to produce a very large predictive sample for returns. We have already noted that this process will add additional kurtosis to the predicted return distribution. We do the same for implied leverages given discretionary wealth uncertainty, so as to produce an independent prediction distribution for the risk-aversion coefficient. These return and appropriate risk-aversion scenarios are combined to form a more comprehensive joint probability distribution, which in general will tend to increase the importance of higher moments for consequent log returns on discretionary wealth. Finally, we optimize security weights over this very large sample according to our objective function, in this case expected log return on discretionary wealth, taking into account any constraints. These may include not only a budget constraint but a prohibition against short positions, as well as many other constraints imposed by the investor's or the investment manager's situation. Dimensionality of input information is thus preserved until the final ranking.

Within this general framework, one may choose from a variety of likelihood models, of greater or lesser complexity, for returns and implied leverage. Derivative securities may be included. With the advent of stronger optimizing algorithms, we can also cope with more complex solution constraints than were addressable when Markowitz first formulated the portfolio optimization problem.

In essence, we are employing a form of the old technique of scenario analysis, as in Markowitz and Perold (1981), but using the calculation speed of the computer to range over many thousands of scenarios.

3.3 Test Examples

Much of the advantage of Bayesian investing can come from more accurate derivation of point estimates for means and covariances assembled through the methods noted in Section 3.1. However, since these effects are already relatively well known, we focus here on an illustration of the remaining improvements that can be obtained through preserving full probability

---

[19] Though log-normal or Student's t distributions are common in the literature, neither these nor a normal return process is a precise model for realistic security return generation. But a multivariate normal likelihood distribution, Winsorized so that returns cannot be worse than −99%, will be used in Section 3.3 in our test examples because we want a best case for Markowitz mean-variance optimization before comparing it with our fully Bayesian model.

[20] We omit estimation of the input parameter posterior distributions because this topic is already well covered in the investment literature. If this estimation is done for each parameter through formal Bayesian methods, a posterior probability distribution will have been created, even if the original purpose was to obtain a point estimate from it. We take a two parameter case as typical, although one can include higher moments at the cost of increased complexity, as along the lines of Harvey *et al* (2002).

distributions in our calculations of best mean log return on discretionary wealth. The examples, which extend the example of Section 2.4, are contrived so that they are small enough to be easily simulated without MCMC, but contain the principles that lead to distortion and inaccuracy when attempting to maximize expected log return on discretionary wealth using a conventional Markowitz mean-variance optimization.

We observe two distinct benefits. First, we find that the asset weights learned from limited samples using log returns on discretionary wealth are more robust than mean-variance optimizations when they are tested out of sample with additional observations from the same likelihood process. A second benefit in accuracy occurs even when our input distributions are fully informed with respect to the likelihood process.

3.3.1 *Test structure*

There is an overwhelmingly rich potential for exploration of examples. We confine ourselves to two test sets. One explores robustness with respect to small-sample error (Table 2) and the other explores accuracy given full information (Table 3). Within each set, we have sub-tests – one constrained to only long positions, and the other permitting short positions as well. Each test compares conventional Markowitz mean-variance optimization to a fully Bayesian discretionary wealth model. The returns used in the tests are structured so as to not to unduly tax Markowitz mean-variance optimization with return higher moments. We include only four assets: cash, bond, and two stocks.

Scenario returns are constructed as follows. For each scenario, asset return means are drawn from a posterior Student's *t* distribution with 5 degrees of freedom, and means of 2%, 3%, 6.5% and 6%. These means are known with standard deviations calibrated at 0.1%, 0.5%, 1.5% and 1.5%. Asset return covariances are drawn from an inverse Wishart distribution with 8 degrees of freedom, centered around covariances based on standard deviations of 1%, 5%, 20% and 20%, and with zero intercorrelations, except that the two stocks have return correlations of 0.99. Finally, for each scenario, the means and covariances drawn are used to generate a set of four normally distributed returns, Winsorized so as to prevent are returns of less than −99%.

A probability distribution for remaining lifetime for our 55-year old woman appears in Figure 1. Note the transformation the shape of the probability distribution takes when mapped into leverage *L*, as in Figure 2. The skew has been reversed as longer lifetimes translate to increasingly larger leverages and consequently greater appropriate risk aversion.

<INSERT FIGURES 1 AND 2 AROUND HERE>

3.3.2 *Procedure and results*

We begin by illustrating Markowitz mean-variance optimization's potential for sensitivity to estimation error; then we demonstrate the impact on Markowitz mean-variance optimization of non-affine transformations and loss of the higher moments of the downstream distributions.

For our robustness test, we first construct 1,000 samples; each contains 100 joint observations of implied leverages and asset returns. For mean-variance optimization, optimal

weights are based on mean sample leverage together with sample return means and covariances formed within each sample. For optimization of log return on discretionary wealth, optimal weights are formed based on maximizing the mean of the logarithmic leveraged returns within each sample.

Each sample's resulting set of weights forms a "rule." The rules are tested on a much larger sample of 900,000 observations. The mean of log leveraged portfolio return is collected for each rule, as well as the variance; the latter may help assess speed of convergence toward best median long-term return. A single-period median is also recorded. The out-of-sample results are shown in Table 2. Note that the log returns shown are for discretionary wealth. They must be converted to arithmetic returns and divided by leverage to measure returns on the investment portfolio.[21]

<center>**<INSERT TABLE 2 ABOUT HERE>**</center>

In the long-only case, the mean deficit experienced by the Markowitz approach as compared to the fully Bayesian discretionary wealth approach was 110 basis points. Although the median deficit was only 20 basis points, that measure overlooks some of the quite poor results at the bottom of the distribution. One explanation is that the Markowitz models did not learn to sufficiently avoid instances in which the combination of longer lifetime and poor investment results was present.

When short positions were allowed, the results were a greatly increased degradation of relative performance for Markowitz mean-variance optimization. The median deficit was 147 basis points and the mean deficit was 334 basis points. This is consistent with Markowitz's Presidential address to the American Finance Association (Markowitz, 1983) warning against excessive short positions in the portfolio. To the extent that reliability is an issue because, after all, we do not face an infinite number of periods, standard deviations are shown as well, and there is greater variation in mean log returns on discretionary wealth among rules learned by conventional Markowitz methods.

Our accuracy test is based on a single, larger learning sample of one million observations, essentially full information. In it, we see that the Markowitz-optimized mean log return on discretionary wealth in the long-only case was only 27 basis points worse than that of the fully-Bayesian discretionary wealth approach. However, when short sales were allowed, even though the Markowitz approach had essentially full information as to the distribution of both returns and leverages, it fared 129 basis points worse.

Table 3 also gives the asset allocation selected by the two approaches. In the long only case, the Bayesian discretionary wealth approach was to take less risk, consistent with the nonlinear effects stemming from products of the long tail on lifespan and thus on leverage. In the case allowing short positions, the Bayesian discretionary wealth approach made less use of short positions. Finally, as in Table 2, it exhibited less variation in log returns.

---

[21] Although a basis point (one hundredth of a percent) does not strictly apply to logarithms, it is convenient terminology in the absence of better.

<INSERT TABLE 3 ABOUT HERE>

3.3.3 *Further discussion*

By varying the conditions of the comparisons between the two approaches, it would be possible to analyze in greater detail the factors contributing to degradation of mean-variance optimized performance when compared against the standard of optimal mean log return on discretionary wealth achieved through a Bayesian approach. Space does not permit an extensive exploration here. However, it is generally apparent that household investors with ample, relatively saving and spending plans, and seeking only broad asset allocation guidance with no consideration of taking on short positions, will not gain much from the Bayesian model.

On the other hand, the benefits of the full Bayesian analysis can be material, with even very small numbers of assets, for two quite different investors. One is the household investor of proportionately lower discretionary wealth and consequently high implied leverage. For such investors, uncertainty with regard to future cash flows, especially because of longevity risk, is likely to create a strongly skewed distribution of possible implied leverages, as well as making higher return moments of greater relative interest because of greater volatility of investment returns on discretionary wealth. Another investor who may benefit from a full analysis is the professional investment manager who may have high leverage, both implicit and explicit, who may take on short positions, and who may use derivatives with strong high moment effects.

4. CONCLUSIONS

Taking advantage of the pioneering efforts of others, we have sought to distill and frame the investment policy problem as simply as possible while retaining sufficient content for a rich variety of applications. The theorist may ask "Why didn't you use utility models more clearly, and where are the stochastic differential equations?" We do not think these are necessary to achieving substantial improvement in understanding investment policy. The practitioner may more justly complain that by trying to do too much, we have provided only a small taste of each suggested application, and that there should be more examples and tests for each. We hope that others will help us fill these gaps.

We should not burden the reader with a recapitulation of the models or all their applications, but we can highlight some of the main points. The models flow naturally from a paradigm shift built on four ideas. First, an augmented balance sheet encompassing the present values of future savings and spending plans to quantify discretionary wealth provides a superior description of position. Second, long-run maximization of median wealth without interim shortfalls is a worthy goal, and it can best be attained in the current period by maximizing expected growth of discretionary wealth so as to provide the normative risk aversion missing from Markowitz mean-variance optimization, so as to allow consideration of higher return moments, and so as to increase opportunities for improving saving and spending plans. Third, adoption of Bayesian logic using full probability distributions to preserve all relevant information overcomes the fragility of point-estimate-based optimizing methods and reduces the need for *ad hoc* constraints to produce usable results. Fourth, effectively integrating uncertainty

31

in saving and spending plans is comparable in importance to dealing with return risks in setting investment policy, and this may be done straightforwardly through a fully Bayesian approach.

We find that conventional Markowitz mean-variance optimization, provided it uses appropriate risk aversion as outlined here, does a good job in many circumstances. We have outlined the scenarios under which this will not be true. In Section 2, we saw that this was when higher return moments were important, most clearly when derivative products are involved. However, more subtly, the importance of higher moments is not simply a function of the size of skewness or kurtosis, and can be dramatic even with relatively conventional securities under certain conditions. The key index of this is the product of the implied leverage, the standard deviation of returns, and what is left after taxation. This is because the impact on expected log return of discretionary wealth from third and fourth moments is roughly proportional to increasing powers of this index multiplied by skewness and kurtosis. Also, because the first two moments of security returns are approximately linear with length of period, but higher moments are nonlinear, an important control over the effect of higher moments is the periodicity with which investment decisions are made within the investment policy. The longer the period between decisions, the more important it is to take account of higher moments.

In Section 3, we found also that conventional mean-variance optimization will be additionally problematic when saving and spending plans are imprecise, and consequently implied leverages to discretionary wealth are uncertain, especially if skewed with a fat-tail toward greater values. We also found that conventional optimization is inadequate as compared to a fully Bayesian approach when the asset allocation contemplated allows short-sales and involves interrelated returns with covariance matrices approaching singularity. This is revealed not only in poor robustness to input error but also in less accuracy. The latter is true even in benign return environments with full information as to the probability distributions from which returns and implied leverages are drawn.

There are two by-products to this investigation we believe are worthy of additional note. First, if we assume that investors implicitly following our suggested normative model actually do amass greater median wealth, they should have a disproportionate impact on pricing. This observation generates a number of descriptive hypotheses predicting and explaining what have been regarded as anomalies. Additional market volatility and pricing of higher moments are included, bearing not only on value and small stock premia, but also on the pricing of out-of the money options and of lock-in provisions. This model also predicts investor clienteles for buying and selling higher moments based on differences among investors in their implied leverages and tax rates.

Second, this work clarified, at least for us, the power of the model to help determine through optimal leverage calculations not only investment decisions but decisions as to saving and spending. Since we have eschewed devices such as time preference parameters to link consumption and investment returns, we cannot say in general whether it is consumption plans or investment plans that should be adjusted. However, it is clear that in any particular instance the costs of each adjustment can be assessed, the resulting expected growth rate can be calculated, and we have at least specified what the product of the two adjustments should look like.

Finally, we recognize that effective investment policy is a solution to a dynamic control problem that we have only approximated. We have seen enough, however, of the contingent nature of the solution to observe that the probability of shortfalls such as experienced recently by many investors in the 2007-2009 period can be substantially reduced. We also conclude that the age-based life-cycle products now being marketed are far from sufficiently customized and flexible to make them good solutions for many investors.

REFERENCES

Black, F. and A. F. Perold (1992) Theory of constant proportion portfolio insurance. *Journal of Economic Dynamics and Control* 16 (3-4): 403-426.

Black, F. and R. Litterman (1992) Global portfolio optimization. *Financial Analysts Journal* 48(5): 28-43.

Booth, D. and E. Fama (1992) Diversification returns and asset contributions. *Financial Analysts Journal*, May/June: 26-32.

Breiman, L. (1961) Optimal gambling systems for favorable games. *Proceedings of the Fourth. Berkeley Symposium on Mathematical Statistics*, 1: 65-78.

Brunel, Jean (2006) How sub-optimal - if at all - is goal-based asset allocation? *Journal of Wealth Management* Fall: 19-34.

Campbell, J. Y. (2006) Household finance. *Journal of Finance* 61(4): 1553-1604.

Carlin, S. and B. P. Chib (1995) Bayesian model choice via Markov chain Monte Carlo methods. *Journal of Royal Statistical Society Series B* 57 (3): 473-484.

Das, S. and H. Markowitz, J. Scheid, and M. Statman (2009) Portfolio optimization with mental accounts. *Journal of Financial and Quantitative Analysis* (forthcoming).

Dybvig, P.H. (1995) Duesenberry's ratcheting of consumption: optimal dynamic consumption and investment given intolerance for any decline in standard of living. *Review of Economic Studies* 62: 287-313.

Fabozzi, F.J., P.N. Kolm, D. Pachamanova, and S. M. Focardi (2007) *Robust Portfolio Optimization and Management*. New Jersey: John Wiley & Sons.

Fama, E. F. (1968) Risk, return and equilibrium: some clarifying comments. *Journal of Finance*, 23(1): 29-40.

Fama, E. and K. French (1992) The cross section of expected stock returns. *Journal of Finance* 47(2): 427–465.

Feller, W. (1968) *An Introduction to Probability Theory and Its Applications*, *Vol. 1, 3rd Edition*. New York: John Wiley & Sons.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin, D.B. (2004) *Bayesian Data Analysis, 2nd Edition*. London: Chapman and Hall.

George, E. and R. McCulloch (1993) Variable selection via Gibbs sampling. *Journal of the American Statistician Association* 88(423): 881-889.

Green, P. (1995) Reversible jump MCMC computation and Bayesian model determination. *Biometrika* 82(4): 711-732.

Grossman, S.J. and G. Laroque (1990) Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods. *Econometrica* 58: 25-52.

Greyserman, A., D.H. Jones, and W.E. Strawderman (2006) Portfolio selection using hierarchical Bayesian analysis and MCMC methods. *Journal of Banking and Finance* 30(2): 669-678.

Hakansson, N. (1971) Capital growth and the mean-variance approach to portfolio selection. *Journal of Financial and Quantitative Analysis* 6(1): 517-557.

Hakansson, N. (1974) Comment on Merton and Samuelson. *Journal of Financial Economics* 1 (1): 95.

Hardin, G. (1968) The tragedy of the commons. *Science* 162 (December): 1243-1248.

Harvey, C. R. and A. Siddique (2000) Conditional skewness in asset pricing tests. *Journal of Finance*, 55(3): 1263-1295.

Harvey, C. R., J. C. Liechty, M.W. Liechty, and P.Müller (2002) Portfolio selection with higher moments. Working Paper, Duke University Durham.

Harvey, C.R., J.C. Liechty, and M. Liechty (2008) Bayes vs resampling: A rematch. *Journal of Investment Management* 6(1): 1-17.

Ibbotson, R.G., M.A. Milevsky, P. Chen, and K. Zhu (2007) *Lifetime Financial Advice: Human Capital, Asset Allocation, and Insurance*. Charlottesville, Virginia: Research Foundation of CFA Institute.

Jacobs, B.I., K.N. Levy, and H.M. Markowitz (2005) Portfolio optimization with factors, scenarios, and realistic short positions. *Operations Research* 53(4): 586–599.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21: 279–292.

Kan, R. and G. Zhou (2007) Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42(3): 621-656.

Kelly, J. (1956) A new interpretation of information rate. *Bell System Technical Journal 35*: 917-926.

LeDoit, O. and M. Wolf (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88(2): 365-411.

Leland, H.E. (1999) Beyond mean-variance: Risk and performance in a nonsymmetrical world. *Financial Analysts Journal* (January-February): 27-36.

Levy, H., and H. M. Markowitz (1979) Approximating expected utility by a function of mean and variance. *American Economic Review* 69: 308-317.

Liu, J.S. (2008) *Monte Carlo Strategies in Scientific Computing*. New York: Springer.

Lunn, D.J., A. Thomas, N. Best, and D. Spiegelhalte (2000) WinBUGS a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10: 325-337.

MacLean, L. C. and W.T. Ziemba. (2006) Capital growth: Theory and practice. In S. A. Zenios and W. T. Ziemba (eds), *Handbook of Asset and Liability Management*, *Vol 1.Theory and Methodology.* Elsevier, B.V., 429-469.

Markowitz, H.M. (1976) Investment for the long run: New evidence for an old rule. *Journal of Finance* 31(5): 1273-1286.

Markowitz, H.M. (1983) Nonnegative or not nonnegative: a question about CAPMs. *Journal of Finance* 38(2): 283-295.

Markowitz, H.M. and A. F. Perold (1981) Portfolio analysis with factors and scenarios. *Journal of Finance* 36(4): 871-877.

Markowitz, H.M. and N. Usmen (2003) Resampled frontiers versus diffuse Bayes: An experiment. *Journal of Investment Management* 1(4): 9-25.

Mehra, R. (ed) (2007) *Handbook of the Equity Risk Premium*. Elsevier Science.

Merton, R.C. (1990) *Continuous-Time Finance*. Cambridge, MA.: Blackwell Publishers.

Merton, R.C., and P. A. Samuelson (1974) Fallacy of the log-normal approximation to optimal portfolio decision making over many periods. *Journal of Financial Economics* 1(1): 67-94.

Meucci, A. (2005) *Risk and Asset Allocation*. New York: Springer Finance.

Michaud, R.O. (2001). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. New York: Oxford University Press.

Perold, A.F. and W.F. Sharpe (1995) Dynamic strategies for asset allocation. *Financial Analysts Journal* 51(1):149-160.

Rachev, S.T., S.J. Hsu, B. Bagasheva, and F. J. Fabozzi (2008) *Bayesian Methods in Finance*. New York: John Wiley & Sons.

Robert, C.P. and G. Casella (2005) *Monte Carlo Statistical Methods. 2nd* .New York: Springer.

Roy, A., (1952) Safety first and the holding of assets. *Econometrica* 20: 431-49.

Rubinstein, M. (1976) The strong case for the generalized logarithmic utility model as the premier model of financial markets. *Journal of Finance* 31(2):551-571.

Samuelson, P.A. (1937) A note on measurement of utility. *Review of Economic Studies* 4(2): 155-161.

Samuelson, P. A. (1990) Asset allocation could be dangerous to your health: Pitfalls in across-time diversification. *Journal of Portfolio Management* 16(3): 5-8.

Sharpe, W. F. (1966) Mutual fund performance. *Journal of Business* 39 (January): 119–38.

Sharpe, W. F. (1994) The Sharpe ratio. *Journal of Portfolio Management* 21(1): 49–59.

Shiller, R. J. (2000) *Irrational Exuberance*. Princeton, NJ.:Princeton University Press.

Stutzer, M. (2000) A portfolio performance index. *Financial Analysts Journal* 56(3): 52-61.

Stutzer, M. (2003) Portfolio choice with endogenous utility: A large deviations approach. *Journal of Econometrics* 116(1-2): 365-386.

Thorp, E. O. (1971) Portfolio choice and the Kelly criterion. *Business and Economics Statistics*. Section of Proceedings of the American Statistical Association: 215-224.

Thorp, E. O. (2006) The Kelly criterion in blackjack, sports betting, and the stock market. In S. A. Zenios and W. T. Ziemba (eds), *Handbook of Asset and Liability Management*, *Vol 1.Theory and Methodology.* Elsevier, B.V., 385-428.

Tobin, J. (1958) Liquidity preference as behavior toward risk. *Review of Economic Studies* (February): 65–86.

Tribble,S.D. and A. B. Owen (2008). Construction of weakly CUD sequences for MCMC sampling. *Electronic Journal of Statistics* 2: 634–660.

Wilcox, J. (2003) Harry Markowitz and the discretionary wealth hypothesis. *Journal of Portfolio Management* (Spring): 58-65.

Wilcox, J. (2008) The impact of uncertain commitments. *Journal of Wealth Management* 11(3): 40-47.

Williams, J. B. (1936) Speculation and the carryover. *Quarterly Journal of Economics 50*(3): 436-455.

Ziemba, W.T. and R.G. Vickson (1975) *Stochastic Optimization Models in Finance*. New York: Academic Press.

TABLE 1. EXAMPLE OF AN EXTENDED BALANCE SHEET
(Millions of dollars)

Investment Portfolio	10.0	Investment Debt	2.2
Implied Assets	1.5	Implied Liabilities	5.3
		Discretionary Wealth	4.0
TOTAL	11.5	TOTAL	11.5
Implied Leverage	2.50		

TABLE 2. ROBUSTNESS OUT OF SAMPLE
(For 1000 Learning Samples of 100 Observations)

Example	Treatment	TESTED MEAN LOG RETURN ON D.W. Cross-Learning Sample Statistics		
		Mean	Median	Std. Deviation
1	Markowitz mean-variance	0.0573	0.0751	0.4313
	Bayesian mean log on DW	0.0684	0.0772	0.3542
2	Markowitz mean-variance	0.0053	0.0418	0.8484
	Bayesian mean log on DW	0.0387	0.0565	0.6761

KEY:

Example 1 -- Short positions not allowed.

Example 2 -- Short positions allowed.

Note: The test sample was 900,000 observations.

TABLE 3. ACCURACY GIVEN FULL INFORMATION
(For a Sample of 1,000,000 Observations)

Example	Treatment	WEIGHTS Cash	Bond	Stock 1	Stock 2	FITTED LOG RETURN ON D.W. Mean	Std. Dev.
1	Markowitz mean-variance	0	0.633	0.367	0	0.08106	0.3199
	Bayesian mean log on DW	0	0.729	0.274	0	0.08377	0.2391
2	Markowitz mean-variance	-0.827	1.430	1.931	-1.534	0.07769	0.5726
	Bayesian mean log on DW	-0.213	0.962	1.322	-1.069	0.09058	0.3084

KEY:

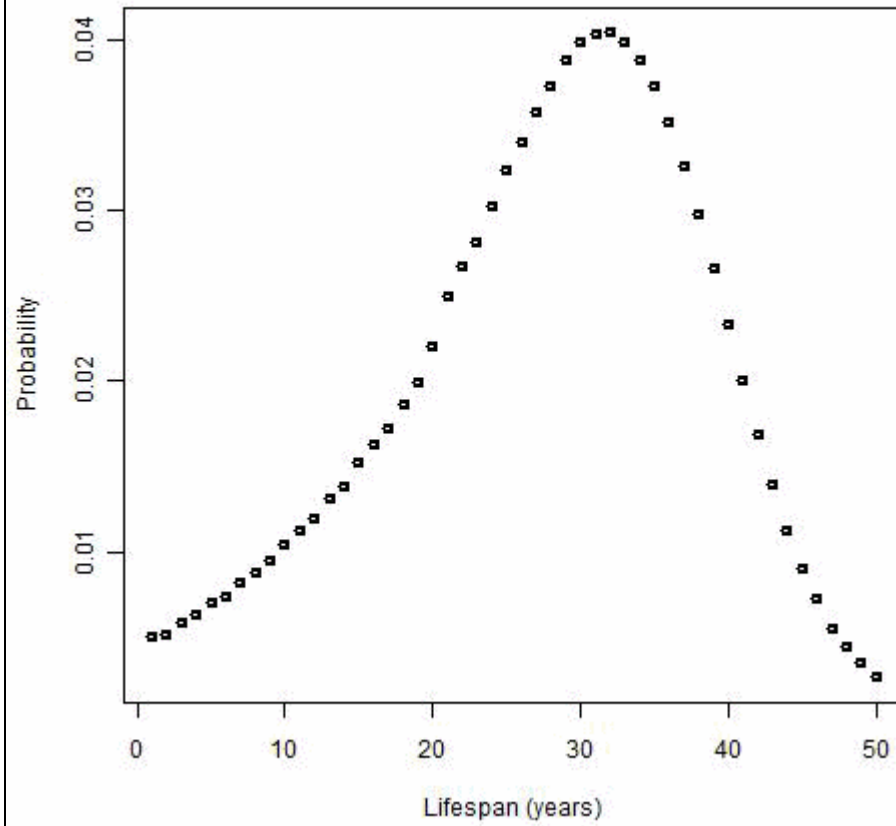Example 1 -- Short positions not allowed.

Example 2 -- Short positions allowed.

Figure 1. Lifespan Distribution

Figure 2. Leverage Distribution