# IS THE EFFICIENT FRONTIER EFFICIENT?

WILLIAM C. SCHEEL, WILLIAM J. BLATCHER,
GERALD S. KIRSCHNER, AND JOHN J. DENMAN*

*Abstract*

*The paper defines plausible ways to measure sampling error within efficient frontiers, particularly when they are derived using dynamic financial analysis (DFA). The properties of an efficient surface are measured both using historical segments of data and using bootstrap samples. The surface was found to be diverse, and the composition of asset portfolios for points on the efficient surface was highly variable.*

*The paper traces performance of on-frontier and off-frontier investment portfolios for different historical periods. There was no clear cut superiority to the on-frontier set of portfolios, although lower risk-return on-frontier portfolios were generally found to perform better relative to comparable, off-frontier portfolios than those at higher risk levels. It is questionable whether practical deployment of optimization methods can occur in the presence of both high sampling error and the relatively inconsistent historical performance of on-frontier portfolios.*

*The implications of this paper for DFA usage of efficient frontiers is that sampling error may degrade the ability to effectively distinguish optimal and non-optimal points in risk-return space. The analyst should be cautious regarding the likelihood that points on an efficient frontier are operationally superior choices within that*

*William C. Scheel, Ph.D., DFA Technologies, LLC; William J. Blatcher, AEGIS Insurance Services, Inc.; Gerald S. Kirschner, FCAS, MAAA, Classic Solutions Risk Management, Inc.; and John J. Denman, AEGIS Insurance Services, Inc.

*space. There are many possible frontiers that optimally fit different empirical samples. Sampling error among them could cause the frontiers to traverse different regions within risk-return space, perhaps at points that are disparate in a decision sense. What is an efficient point on one frontier may be inefficient when calculated from a different sample. The paper finds the use of an efficient surface to be helpful in diagnosing the effects of such sampling error.*

ACKNOWLEDGEMENT

The authors gratefully acknowledge the very constructive comments of reviewers of the paper.

## 1. INTRODUCTION

Companies choose among investments often with the purpose of optimizing some goal and always limited by constraints. Assets are divided among competing investment alternatives with the hope that risk will be minimized for a desired level of return, either investment return or overall return. When the allocation fulfills the goals within the boundaries of constraints, it is thought to be efficient. The allocation is deemed to be a member of the efficient set at a point on an efficient frontier. It is efficient because it dominates off-frontier, interior points in the risk-return space.

This paper investigates this popular investment allocation strategy in two ways. First, it seeks to determine what the sensitivity of the frontier is to possible sampling error in risk-return space. Second, both on-frontier and off-frontier portfolio allocations for actual series of returns are tracked for their respective performance. We begin with an apologue; it gives the reader both a rationale and definition of what we mean by an efficient surface.

### *1.1. A Sampling Error Apologue*

I walk into a casino with shaky knees and a rather small stake. Betting doesn't come easily for me, and I expect to lose the stake. Ralph told me I would lose it. But I have a bevy of information gleaned from experiments Ralph did with a computerized simulation of a craps table. One of the items I call "knowledge" is the efficient surface he made for me. Ralph said it would help me understand the risk-return properties of the craps table and guide me in allocating my stake among the various bets that I can make.

"There are many bets you can make at the table," Ralph explained, "'Come,' 'Big-8' and lots of others. I think of the gaming as a multivariate process. Of course, it has probabilities that are objective and can be measured. Do you want me to figure out the combinatorics of the craps game and derive analytic solutions for optimal bet placement? My consulting fee might be a bit high because the math will take awhile, but I could do it."

I mentally recalculated my meager stake and replied, "Is there a less expensive way?"

Ralph shrugged and said, "Sure. I can use a computer simulation I have and take a sample of game outcomes. I'll use the sample to empirically develop a covariance matrix for some of the bets. Then, I'll figure out which combinations of bets have minimum variance for a particular payoff. You can choose which risk-return profile of bets is best for you. You'll be able to allocate your stake more efficiently. By the way, this is called an efficient frontier—it gives a profile of bets that are expected to produce a given return with minimum variance. I'll do a sample of 25 games each with a combination of various bets. This will keep the cost down."

"Well, okay," I replied, "but will this single efficient frontier really work?"

"What do you mean, 'single frontier'?" he asked.

"What if the sample your computer simulation comes up with is unusual?" Ralph scratched his head, and I continued, "You measure this thing you call a sample covariance matrix. But what if you took a different sample? You'd get a different sample covariance matrix, right?"

"Yes."

"And it might be different?"

"Yes. Even materially different."

"So, your efficient frontier (EF) is subject to sampling error—it was empirically derived from the sample of only 25 games." I then asked, "What if you had a second sample of 25 games and did another mathematical optimization. So we now have 2 different EFs; both do the same thing, but the answers are different. Which one do I use when I walk into the casino?"

Ralph exclaimed, "I'll take a sample, and then another, and another. Each will have a different EF. Then, I'll plot each point of the samples' EFs in risk-return space. I'll count the number of times the various EFs traverse a particular cell in that space. Maybe 10 EFs traverse the cell at the coordinates (10,15). Maybe only 3 EFs traverse the cell at (1,3). Don't you see? Just by counting the number of times the sample EFs traverse a region in risk-return space and normalizing the count to probabilities, I can measure an efficient surface."

I asked, "Why is the surface important?"

Ralph was now animated. He leaped to his feet. "Because, if the various sample EFs all traversed the same cells, the EFs would all be the same—there would no sampling error. What if the surface is spread out? Suppose some sectors of it are relatively flat? Then the efficiency of the EFs varies. Would you prefer to pick a point on the surface (with a particular combination of bets) that appears most often among different EFs? Probably you would. You want the surface to be tightly peaked.

In three dimensions, that's a ridge or very pointy hill; in two dimensions, it is a probability distribution with little variance."

He then went home to begin the chore of sampling and constructing an efficient surface for me. I began to think, "A single efficient frontier is measured from data. We often think of the data being a sample from a replicable experiment. If a sample of dice games is observed, the *n*-tuple bet outcomes for the correlated bets are the empirical data source for an optimization. It is easy to see how different samples can be drawn when talking about dice games. But the world of security returns is different from a craps table. What is a sample there? What is the meaning of sampling error, and how might it affect the way I measure efficient frontiers? Would an EF for securities really be efficient?"

These are important questions—ones addressed in this paper. It is difficult to think of how we'd repeat an experiment involving security returns. Is a series of experiments one that uses different historical periods of returns? Is it a bootstrap of a broad segment of history? These are the two approaches that are equivalent to sampling and measuring sampling error. The result of our measurements is an efficient surface.

### *1.2. Roadmap for the Paper*

Section 2 of the paper lays the groundwork for measuring sampling error that affects efficient frontier measurement. We examine two approaches that seem particularly useful for dynamic financial analysis. We also review the literature relating to EF efficiency. Section 3 introduces the notion of an *efficient surface*—this is a construct for understanding and measuring sampling error in EFs. In this section we describe the methodology and data set used in our study.

The main body of results is presented in Sections 4, 5 and 6. We measure forecast performance of efficient frontiers in Section 4. We are particularly concerned about the performance of off-frontier portfolios. Are they really inefficient? Do on-frontier portfolios dominate performance, as we might anticipate given that they are billed as "efficient"? The evidence we present in Section 4 shows instability in EFs derived both with historical segments and bootstrap samples. This leads us to conclude later that caution should be exercised when using efficient frontiers in DFA analysis.

On the road to this conclusion, we closely examine the efficient surface in Section 5. It portrays sampling error from two different perspectives—historical and bootstrap sampling. The efficient surface is a useful construct for visualizing sampling error in EFs. We observe that such error is particularly large in the high risk-return regions of the surface. This observation is reinforced in Section 6 by observing the diversity of portfolio composition as we compare different historical segments.

The final section is devoted to conclusions and cautions on the use of EFs in DFA work. We conclude that EFs may not warrant the term *efficient.* Their best use may be as advisory measurements concerning the properties of risk-return space.

## 2. SCENARIO GENERATION IN DFA

Dynamic financial analysis involves scenario generation. There are many types of scenarios that are simulated so that the model builder can measure a hypothetical state-of-the-world with accounting metrics. Asset generators typically create returns for invested assets. They model exogenous economic conditions. Each modeler sees the forces of the financial markets unfolding according to a set of rules. The rule set is almost as diverse as the number of modelers.

Some DFA model builders prefer stochastic differential equations with various degrees of functional interrelatedness. The transition of returns over time, as well as the correlations among different asset components, always are represented in multiple simultaneous equations. Other DFA modelers use multivariate Normal models, which conjecture a covariance matrix of investment returns. These models do not have time-dependent transition modeling information. Such an efficient frontier, by definition, has no time transition properties. A sample taken from *any* sub-period within the time series would contain sampling error, but otherwise, the investment allocation would be unaffected.

Both approaches begin with a single instance of reality. They both purport to model it. One approach, stochastic equations, uses largely subjective methods to parameterize the process.[1] Another approach to modeling clings to assumptions that seem to be or are taken to be realistic.[2] Both produce scenarios that are deemed sufficiently similar to reality to represent it for the purpose at hand.

The efficient frontier calculation can be a constrained optimization either based on a sample from a historic series of returns or a derived series with smoothing or other ad hoc adjustment. Alternatively, an EF may be created from simulated DFA results. Both the efficient frontier and DFA asset-based modeling are using the same set of beliefs regarding the manner by which statistically acceptable parameters are used.[3] They both start with a single historic time series of returns for various component assets.

### *2.1. Two Viewpoints on the Use of Efficient Frontiers*

The practitioner has a straightforward objective: define investment allocation strategy going forward. Today's portfolio allocation leads to tomorrow's result. The portfolio is then rebalanced relative to expectations. The new one leads to new results. The

[1]The calibration may depend on examination of stylistic facts, but there seldom is formalized, statistical hypothesis testing to judge whether the facts can be accepted as such or whether the representation of these facts in the model is really a scientific determination.

[2]Some models use multivariate Normal simulation for rendering investment returns for consecutive periods. There usually is an assumption that the covariance matrix used for multivariate Normal simulation is stationary from period to period in these models.

[3]DFA and optimization do have a critical junction. Some DFA modelers believe they understand time dependencies within period-to-period rates of return. EF attempts to optimize *expected* return. If there is a time dependence conjectured, it should be factored into the expected returns used to build the EF for any period.

cycle repeats. Where does the chicken end and the egg begin? In practice, the practitioner has only one instance of yesterday's reality and tomorrow's expectations from which to construct a portfolio and a model.

There are at least two approaches to using a DFA model to define an investment allocation. In one, a DFA analyst might set up an initial allocation of assets using an efficient frontier obtained from quadratic optimization on a prior historical period. A DFA model would be run repeatedly—a different state-of-the-world would ensue each time, and a different reading obtained for the metric. These simulations produce endpoints in the modeled risk-return space. In this approach, one beginning asset allocation leads to many different observations about endpoints. The reason they are different is that, although each starts with the same state, the model simulates various outcomes. Each hypothetical one probably leads to a different endpoint for the planning horizon.

But, another viewpoint exists.[4] We refer to it as the hybrid approach. Suppose that history serves a valid purpose in calibrating a model but should not be used to define a beginning allocation. In this viewpoint, the investment mix *is suggested by the optimizer*. DFA serves only to measure what could happen with some hypothetical starting allocation.

The optimizer deals the cards in this deck, and DFA traces where the cards lead.[5,6] The optimizer, not the modeler, submits

---

[4]Correnti, et al. review an approach similar to the hybrid model described here.

[5]The optimizer posits a trial solution; it consists of a certain portfolio allocation. This trial allocation does not depend on any prior allocation of assets. Rebalancing that ensues during the optimization period (and under the control of the DFA model) also is unknown to the optimizer. The objective value that is returned by the model is driven by the initial trial solution and model outputs that build on the trial solution.

[6]Investment rates are forecasted by the DFA model, which might use multivariate Normal simulation. There may be an overlap between what the optimizer uses and what the DFA model uses. For example, the covariance matrix used for the multivariate Normal simulation is estimated from historical data and generally is assumed to be stationary during the forecast period. It is used both by the optimizer and by the DFA model.

an initial allocation for review. In this hybrid approach, there is no initial portfolio based on optimization using prior history. In the hybrid model, the optimizer finds a portfolio, which leads to an ex post optimal result. The metric used in this optimization is part of the DFA model—it is calculated by the accounting methodology of the model as it generates future states of the world. It may be difficult to reconcile the use of efficient frontiers for investments within hybrid-DFA modeling that, on the one hand, believes there is a historically dependent component that can be used for calibration, but rejects the use of data to define a starting portfolio. Yet, on the other hand, simulations of that model are derived to construct an efficient frontier. It may appear as though history has been rejected as information for the purposes of decision making, yet *indirectly* it is used to represent the future. The starting portfolio in the hybrid approach is based at least indirectly on modeling and should represent an analyst's expectations. These expectations are in theory built into the model for return scenario generation, and that model was calibrated to history in some fashion.

In DFA work, a performance metric is chosen. This metric is measured within a risk-return space. The metric must be measurable according to the chosen accounting framework. Risk might be variance, semi-variance or some chance-constrained function of the metric. In the real world, the corporate manager is rewarded for favorable performance of the metric and often penalized by unwanted risk in the metric. The volume of investment in various stochastic components affects a metric's performance. The operational question is how should an allocation be made to investments so that performance of the metric is optimized.

In the forecast period, the modeler generates a scenario of unfolding rates of return using, say, a multivariate, time-dependent asset model. An example would be any of the multi-factor mean reversion models in use today. The simulated progression of returns for a scenario generated by one of these models is affected by an underlying mechanism that forces unusual deviations in the path back towards an expected trajectory of returns. The DFA model typically ties in some way the business operations to the simulated economic environment.[7] This economic scenario typically generates other economic rates, such as the rate of inflation. A scenario that is generated by the economic model is taken to be exogenous; it is mingled with expectations about corporate performance. The company's *operations* are tied to the exogenous influences of the economic scenario.

In the end, this modeling process is repeated many times for the optimizer in the hybrid model. The optimizer requires an answer to the question: Given an initial investment allocation, what is the end-horizon performance of the metric? The optimizer forces the model to measure the result of a simulation experiment given only an initial investment allocation. The model takes the allocation and produces an experimental point in risk-return space. All that is required of the model is its ability to measure the trajectory of the metric within the company's business plan and a beginning allocation of assets. In this regard, the hybrid model is using a sort of dynamic programming approach to optimization. The possible outcomes are considered, and the most desirable traced back to the inputs (initial allocation). The hope is that the optimized feasible set is robust relative to possible stochastic outcomes in the model trajectory. The efficient frontier traces the allocations necessary to achieve various points in this risk-return space. All of this raises the thorny question of subsequent performance dominance of the on-frontier portfolios in the hybrid model. Do EF points truly dominate the performance of off-frontier frontier points—portfolios that are thought to be inefficient and have higher risk for the same return level?

The reason that this is a hybrid approach is that DFA modeling is not deployed on an optimal asset allocation derived directly from the prior time series. Rather, DFA is combined with

[7] A typical behavioral pattern for business growth is modeling it as a function of inflation, which was generated by the economic scenario. Another is to tie severity in claims to underlying inflation as unfolded in the economic model simulation.

optimization to answer the single question: How should the portfolio be immediately rebalanced to achieve an optimal point in risk-return space over the future DFA planning horizon?

Two portfolios can be devised through optimization procedures—one is based on historical results prior to the start of the simulated future time periods. Another one involves allocations that are selected and tried by the optimizer—the DFA model is integral to this second approach. The latter hybrid optimization uses DFA-measured metrics in the optimizer goal function. If applied over the course of the simulated future time periods, and according to the plan of the DFA model, the hybrid approach would seem to yield optimal results at the end of the simulated time horizon. There is no reason to suppose that these two approaches produce the same initial portfolios. Which one is the real optimum?

During the planning horizon, the hybrid model may ignore imperfections that, in real life, might have (and probably would have) been dealt with by ongoing decision making. The EF could have been recalculated with realized data and the portfolio rebalanced. The published state-of-the art in DFA modeling is unclear in this regard; but it may be that no intra-period portfolio optimization is done by DFA models between the time the analysis starts with an allocation posited by the optimizer and when it ends, say, five years later with a DFA-derived metric. It is inconceivable that an organization would mechanically cling to an initial, EF-optimal result for an operational period of this length without retesting the waters.[8]

[8]There is no reason other than a few computational programming complexities why intra-period optimizations cannot be done within DFA models. The question is whether they are, or they are not, being done. For example, the DFA model can simulate a wide variety of rebalancing strategies including the real-life one that involves a rebalancing trigger for simulated portfolios whose allocation has deviated from a recent EF by some amount. Mulvey, et al. [1998, p. 160] describe an $n$-period simulation wherein such rebalancing is triggered. In addition, Mulvey, et al. describe the use of optimization constraints in a clever way to achieve an integration of strategic, long-term optimization with short-term tactical objectives. However, a DFA model that allows intra-period optimization must also capture the transaction and tax costs associated with the intra-period rebalancing and re-

### *2.2. Limitations of this Study for Use of the Efficient Frontier in DFA*

We do not do a complete DFA analysis—there is neither a liability component nor a conventional DFA metric such as economic value of a business enterprise. Rather, the data are limited entirely to marketable, financial assets. Nevertheless, we believe our findings are of value to DFA work. If the efficient frontier produced solely within a traditional investment framework has unstable properties, these instabilities will apply to its use in DFA work were it to be calculated and used in a similar way.

### *2.3. Other Investigations of the Efficacy of EF Analysis*

Michaud has extensively investigated the use of EFs with particular regard to general efficacy for forecasting. For example, he has shown [1998, pp. 115–126] that inclusion of pension liabilities can substantially alter the statistical characteristics of mean-variance (MV) optimization for investment portfolios.

Michaud's book [1998] examines efficient frontiers both with respect to their inherent uncertainty and what might be done to improve their worthiness. He suggests that the effects of sampling error may be improved using a methodology described as a *resampled efficient frontier*. The motivation for some kind of improvement over classical EFs is that "...optimized portfolios are 'error maximized' and often have little, if any, reliable investment value. Indeed, an equally weighted portfolio may often be substantially closer to true MV optimality than an optimized portfolio." [Michaud, 1998, p. 3].

The determination of a resampled efficient frontier is complex; Michaud has patented it. Although his book exposes the core of the method that he believes improves on forecast error, there is no empirical evidence provided in the book that a resampled efficient frontier has this desirable effect. Interested

---

optimization. See Rowland and Conde [1996] regarding the influence of tax policy on optimal portfolios and the desirability of longer term planning horizons.

readers are directed to his book. The concept of an efficient surface espoused in our paper is built on different constructs. We will readdress the important work of Michaud at a later point in the paper. We now turn to the definition and measurement of an efficient surface.

## 3. THE EFFICIENT SURFACE

An efficient frontier consists of points within risk-return space that have minimum risk for a return. If there were a time-stationary, multivariate probability distribution for prior history, then history is a sample from it. History, therefore, would have sampling error.[9]

The concept of a conditional marginal probability distribution either for return or risk emerges, and it, too, would have sampling error. We discuss the properties of this marginal distribution, an equi-return slice of the efficient surface, in Section 5.1.

Were the instance of reality to be a sample, what is the sampling error?

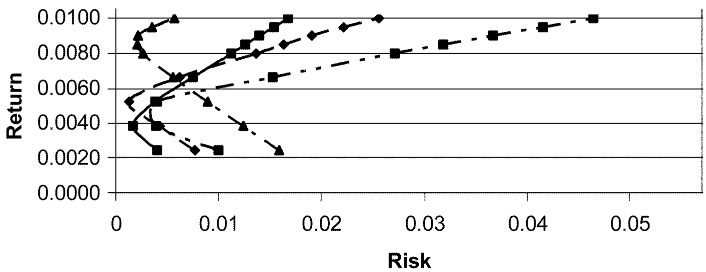
Figure 1 shows efficient frontiers for random 5-year blocks of history. The EFs were derived from monthly returns beginning in January 1988. Each curve in Figure 1 requires optimizations for a 5-year history of returns. The block of monthly returns was picked at random from the entire time series. The points along each EF are obtained from separate passes through the data with the optimizer. On each pass, one of the constraints differs. That constraint is the requirement that the average portfolio return be a specified value in the return domain. The optimizer's objective function is the minimization of variance associated with that portfolio expected return.

---

[9]If there were conjecture, the multivariate distribution would be subjective, and the efficient frontier would be the subjective frontier. A subjectively derived EF has no sampling error, but it may lose *operational* appeal when represented in this manner, because subjectivity requires difficult reconciliation within a corporate, decision-making framework.

FIGURE 1

COMPARISON OF EFFICIENT FRONTIERS FOR DIFFERENT TIME PERIODS




Each EF in Figure 1 consists of nine points; each point involves a separate quadratic optimization. For example, one of the optimization constraints is the portfolio expected return, which is set to an equality condition. There were nine different expected returns used in the study; one was a monthly return of 0.004. An examination of the figure at this value shows a point for each of the four EFs. An empirically derived covariance matrix was determined for each of the four time series illustrated in Figure 1 as well as for hundreds of others that are not shown. The juxtaposition of the EFs displays a tangle of overlapping, crisscrossing curves.[10] This illustration can be viewed as sampling with replacement from a historical sample; it is appropriate, then, to view the figure as illustrative of a probability surface. It is a surface showing the extent of sampling error provided there has been a stationary, multivariate distribution of components' returns.[11] Figure 1 indicates that it may be hazardous to

[10]Some segments of EFs such as those shown in Figure 1 can be indeterminate. This is because the quadratic optimizer could not identify a feasible set of investment alternatives for all of the average returns chosen in the analysis. There is a small probability of overlap of data because the 5-year blocks of returns used for each EF could have overlapping sub-periods of time.

[11]The population distribution is unknown, but it is estimated from the historical record by calculation of an empirical covariance matrix for each historical block.

accept any particular segment of history as the "best estimator." This figure shows only several of the EF curves that build up an efficient surface. Examples of efficient surfaces appear later in Figures 8 and 10. The distribution of risk in a cross-sectional slice of this efficient surface also is reviewed in Section 5.

The positions and slopes of the EFs in Figure 1 are wildly different, and were other historical EFs to be included, the complexity would be greater. This lack of historical stability casts doubt on the operational validity of a particular efficient portfolio actually producing optimal performance. The figure also hints that off-frontier portfolios may perform as well as or better than on-frontier portfolios. We examine this question of forecast reliability in detail in Section 4.

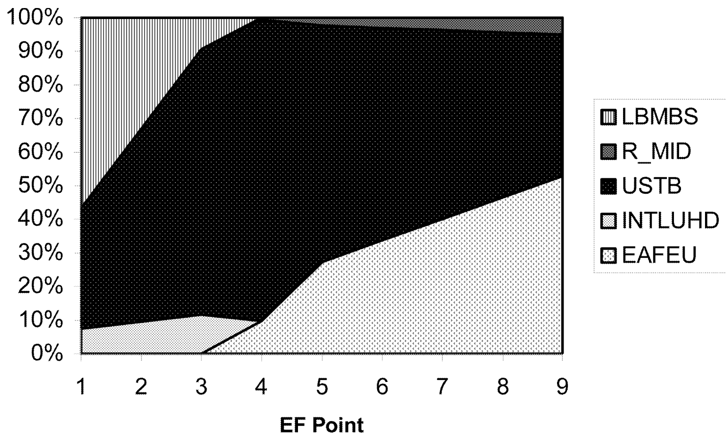
In addition to the positional changes in EFs over time, there is dramatic change in portfolio composition along the curve of each EF in Figure 1. Examples of the change in portfolio composition for EFs appear in Figures 2a and 2b. Each chart is categorical—a tic mark on the $x$-axis is associated with one of nine optimization points. Each chart shows a stacked area rendering of the proportion of an asset component within the efficient set. If the reader views either Figure 2a or 2b from left to right, the unfolding change, and possible collapse, of a particular component is illustrated. This type of chart is a useful way to show a component's contribution to the efficient set moving along the EF from low risk-return to high risk-return portfolios.

There is faint hope that the two different EF portfolio compositions shown in Figures 2a and 2b will operationally produce the same result when put in practice—were this to be a reasonable representation of the effects of sampling error, the operational use of efficient frontiers would be questionable; sampling error swamps operational usefulness and forecast responsiveness.

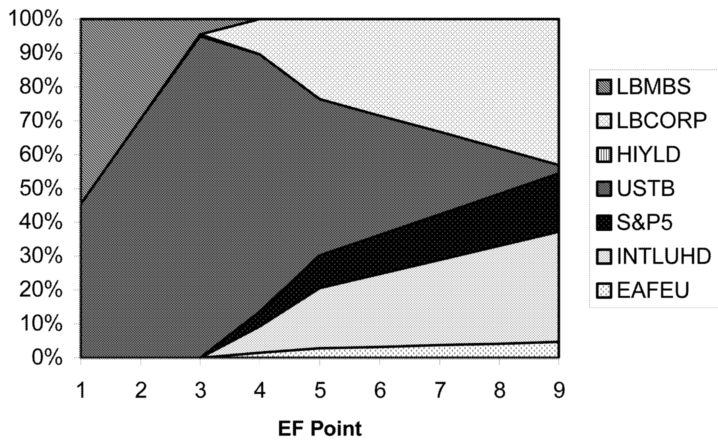
However, another illustration, Figure 3, indicates that if history is a sample from a multivariate distribution, there should be optimism that the efficient frontier evolves slowly, at least measured in monthly metrics. This figure shows EFs calculated

# FIGURE 2

## PORTFOLIO COMPOSITION FOR DIFFERENT EFFICIENT FRONTIERS. (a) COMPOSITION A, (b) COMPOSITION B.



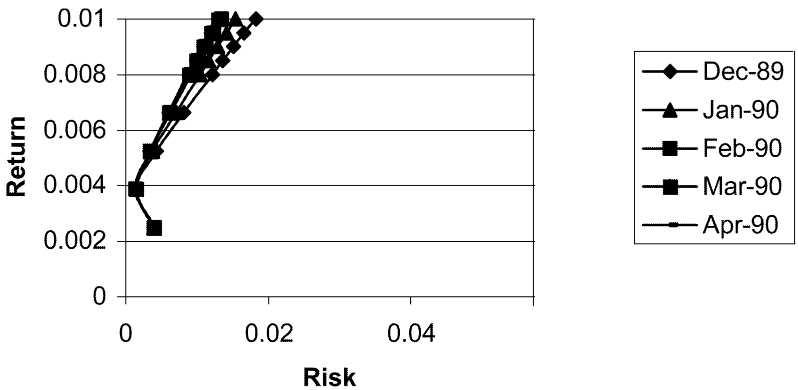

(a)




(b)

FIGURE 3

EFFICIENT FRONTIERS FOR CONSECUTIVE TIME PERIODS




from consecutive, overlapping historical blocks of time. In this case, the time interval between between consecutive EFs is one month. The stability deteriorates fastest at higher risk-return levels. The result was found to hold for a wide variety of consecutive historical blocks starting at various points since 1977. This stability may provide an operational basis for investing in an on-frontier portfolio and seeing its performance prevail over off-frontier portfolios, at least for relatively short planning horizons.

There are other ways to use the historical record. The paper shortly will turn to the use of the bootstrap as a method of measuring sampling error. First, the data and manipulation methods are described in more detail.

### *3.1. Data Manipulation*

This study uses the time series described in Appendix A: Review of Data Sources. Except where gaps were present in the historical record, the portfolio returns are actual.[12,13]

[12]The data represent returns for a selected group of investment components. There was no attempt to filter or smooth the time series in any way. However, a few gaps in the historical record were interpolated.

The data were used in two ways: (1) bootstrap samples were made from the original time series in an attempt to approximate sampling error phenomena, and (2) various historical series of the data were used for performance analysis. The study examines period segmentation and the performance of efficient and inefficient portfolios for different forecast durations.

### *3.1.1. Historical Performance Analysis*

In this section of the paper, data for an efficient frontier are extracted for a historical period and used to evaluate the efficient frontier. The on-frontier portfolios are minimum variance portfolios found using quadratic programming.[14] Off-frontier portfolios also were calculated.[15] The study is concerned with whether the performance of off-frontier portfolios really were inefficient compared to the performance of on-frontier portfolios.

## *3.2. Bootstrap Sampling*

A bootstrap sample of a data set is one with the same number of elements, but random replacement of every element by drawing with replacement from the original set of data. When this process of empirical resampling is repeated many times, the bootstrap samples can be used to estimate parameters for functions of the data. The plug-in principle [Efron and Tibshirani, 1993, p. 35] allows evaluation of complex functional mappings from

---

[13]One technique for deploying efficient frontiers within DFA analysis involves removal of actual values from the data series used in optimization. These points in the actual time series may be deemed abnormalities. The efficient frontier calculation does not use all available data or uses them selectively. See Kirschner [2000] for a discussion of the hazards of historical period segmentation.

[14]All optimization was done using Frontline Systems, Inc. *Premium Solver Plus V3.5* and Microsoft Excel.

[15]It is possible to restate a portfolio optimization problem to produce off-frontier portfolios. These are asset allocations for points in risk-return space that are within the concave region defined by the set of efficient points. They are portfolios with variance greater than the minimum variance points for the same expected returns. They were found by goal equality calculation using the same constraints as were used for minimum variance optimization. However, the equality risk condition was set to a higher level than found on the efficient frontier. Non-linear optimization was used for this purpose, whereas quadratic optimization was used for minimum variance optimization.

examination of the same functional mapping on the bootstrap samples. The function $\theta = t(F)$ of the probability distribution $F$ is estimated by the same function of the empirical distribution $\hat{F}$, $\hat{\theta} = t(\hat{F})$, where the empirical distribution is built up from bootstrap samples. This technique often is deployed for the derivation of errors of the estimate.

The plug-in features of a bootstrap enable inference from sample properties of the distribution of bootstrap samples. The plug-in properties extend to all complex functions of the bootstrap, including standard deviations, means, medians, confidence intervals and any other measurable function. The EF is one of these functions.

The bootstrap is used in this paper to illustrate the impact of sampling error on the EF.[16] EF is a complex function of the historical returns from which it was calculated. If the sample is from a larger, unknown domain, the bootstrap principles apply. In the case of correlated investment returns, a segment of history might be thought of as a sample, but it may not be operationally meaningful because of sampling error. Yet, the use of the historical data in DFA applications treats it as though it were meaningful, representative, and *not* a sample.

The behavior of the EFs for our bootstrap samples is a nonparametric technique used to evaluate the effect of sampling error, were history to be properly thought of as a sample. Because actuarial science is built largely on the precept that past history, even of seemingly unique phenomena, really is a sample, we too proceed along this slippery slope.

### *3.2.1. Bootstrapping* n-*Tuples*

The $n$-tuple observation of correlated observations at time $t$ can be sampled with replacement. This technique was used by

[16]The bootstrap has been used in connection with mean-variance optimization by Michaud and others in an attempt to improve performance of EF portfolios. See Michaud [1998].

Laster [1998]. The experiment is similar to drawing packages of colored gum drops from a production lot. Each package contains a mixture of different colors that are laid out by machinery in some correlated manner. Suppose the lot that has been sampled off the production line contains $n$ packages. A bootstrap sample of the lot also contains $n$ observations. It is obtained by draws, with replacement, from the original *sample* lot. The $n$-tuple of investment returns at time $t$ is analogous to a package within the lot of gum drop samples. The historical sequence of correlated returns is analogous to the mix of different colors of gum drops in a package. The analogy halts because we know the lot of gum drop packages is a sample. We never will know whether the sequence of historical, $n$-tuple investment returns is a sample in a meaningful sense.

The data consist of a matrix of monthly returns; each row is an $n$-tuple of the returns *during a common interval of time* for the component assets (columns of the matrix); the value of $n$ was ten and measures the use of the ten investment categories described in Appendix A: Review of Data Sources. The bootstrap method involves sampling rows of the original data matrix. An $n$-tuple describing the actual returns for asset components at an interval of time is drawn and recorded as an "observation" in the bootstrap sample. Because this $n$-tuple can appear in another draw, the process involves sampling with replacement. This randomized choice of an $n$-tuple is repeated for each observation in the original sample. When the original sample has been replaced by a replacement sampling of the sample, the result is referred to as a bootstrap sample. This process of drawing a bootstrap sample can be repeated many times, usually in excess of 2,000.

Each bootstrap sample has both a measurable covariance matrix and an efficient frontier that can be derived using that covariance matrix. It is unlikely that any two bootstrap samples will necessarily have the same covariance matrix. Each sample can be subjected to mathematical optimization to produce an efficient frontier. The study asks whether this frontier is stable across the samples. Instability is measured in two ways. First, the bootstrapped efficient frontier may fluctuate from sample to sample. This means that the distribution of risk for a return point on the EF is not a degenerate distribution that collapses to a single point. Rather, there is a range of different portfolio risks among the bootstrap samples at a given return. There is a probability distribution associated with risk, given a return among the bootstrap samples. In other words, the study attempts to measure the distribution, and the study views that distribution as a measure of sampling error in risk-return space as it affects the calculation of an efficient frontier.

Second, the portfolio allocations may diverge qualitatively among bootstraps. Were portfolio allocations to be about the same in an arbitrarily small region of risk-return space among different bootstrap samples, the practical effects of sampling error would be small.

### *3.2.2. Extension of the Bootstrap Sample as a DFA Scenario*

The bootstrap samples can be used in the way a DFA model might have used the original historical data, including their direct use within the calculation of the DFA results as a random instance of investment results. They are the source of DFA scenarios. This paper suggests how that direct use of the bootstrap might unfold in a DFA liability-side simulation, but it does not deploy it in that manner.[17,18] The authors have a less ambitious objective of examining just the performance of the efficient frontier built from bootstrapping investment information.

---

[17]Although the $n$-tuple used in this paper is a cross-sectional observation of returns, it can be expanded to a cross-section of the entire business environment at time $t$. This includes all economic aggregates, not just rates of return. Any flow or stock business aggregate that can be measured for interval $t$ is a candidate for the $n$-tuple. This would include inflation, gross domestic product, or any worldly observation of the business climate prevailing at that time. A bootstrap sample can be used as a component of a larger simulation requiring simulation of these worldly events.

[18]DFA model builders spend time modeling empirical estimates of process and parameter risk [Kirschner and Scheel, 1998]. Bootstrapping from the data removes much of this estimation work and leaves the data to speak for themselves.

### *3.3. Sampling Error within Risk-Return Space*

There is no clear-cut method for estimating sampling error that may exist in risk-return space. We do not know the underlying distribution generating a historical sample. We do not know whether a population distribution, were it to exist, is stationary over any time segment. We might, however, view history as an experimental sample, particularly if we want to use it to forecast corporate strategic decisions using DFA.

Sampling error can be envisioned and approximated in different ways for this hypothetical unfolding of reality. One way is to break the actual time series into arbitrary time segments and ask whether a random selection among the subsets of time leads to different, operationally disparate results—these would be EFs based on the sub-segment of time that have portfolio allocations disparate enough to be viewed as operationally dissimilar. If they are dissimilar enough to warrant different treatment, a sampling distribution of interest is the one measured by the effects of these time-period slices.

Another approach is to envision prior history as an instantiation, period-to-period, from an unknown multivariate distribution. The sampling error in this process is driven by a multivariate distribution. Depending on our model, we may or may not place dependencies from prior realizations on this period's realization. That is, for DFA investment return generation and intra-period portfolio rebalancing, the multivariate model may be stationary or non-stationary with respect to time.

#### *3.3.1. Michaud's Efficient Frontier*

Michaud [1998] approaches the measurement of sampling error effects on EF in a different way. Although his approach differs, his overall conclusions are important and consistent with many of our findings. He notes [1998, p. 33], "The operative question is not whether MV optimizations are unstable or unintuitive, but rather, how serious is the problem. Unfortunately for many investment applications, it is very serious indeed." Our paper will draw a similar conclusion.

He does not refer to an efficient surface but calculates a "resampled" portfolio that seems to capture some similar properties. Michaud uses multivariate Normal simulations from the *same* covariance matrix used to calculate EFs. This covariance matrix is from a sample of data—the data observed during some historical period. Just what definition of sampling error has been accommodated in the Michaud resampled portfolio is unclear.

One of the Michaud simulations is not equivalent to a bootstrap sample used in this study. Michaud's approach does not attempt to adjust for a primary source of sampling error—sampling error in the covariance matrix. In our study, each bootstrap sample has an independently measured covariance matrix. Using the DFA jargon of Kirschner and Scheel [1998, pp. 404–408], Michaud's approach may not account for parameter risk in the underlying returns generation mechanism. The ranking mechanism used by Michaud to combine EFs derived from various multivariate Normal simulations may distort risk-return space because each EF is segmented in some non-linear fashion to identify equally ranked points in risk-return space [Michaud, 1998, p. 46, footnote 11]. The portfolio profiles for identically ranked EF points are averaged, yet it is not clear that equi-ranked points fall within the same definition of risk-return space.

### *3.4. Importance to DFA Scenario Generation*

This paper cannot and does not attempt to rationalize the process underlying investment yields over time.[19] Rather, the model builder should be careful to design the DFA model to be in accordance with perceptions about how a sampling methodology may apply. The use of the model will invariably mimic that viewpoint.

[19]What if there were no common observable stationary probability measure for security prices? Kane [1999, p. 174] argues we must use utility measurements.

If, for example, one views history in the fashion imagined by a bootstrap of $n$-tuples, and if that view does observe operational differences, then one can create scenarios from bootstrap samples. No more theory is required. Hypothetical investment returns are just a bootstrap sample of actual history.

Similarly, if EFs for historical periods produce superior performance in forecasting (compared to portfolios constructed from off-frontier portfolios derived from the same data), then the use of an empirically determined covariance model and multivariate Normal simulation makes a great deal of sense.

### *3.5. Importance to DFA Optimization*

Optimization often is used within DFA and cash flow testing models to guide portfolio rebalancing. The DFA model usually grinds through the process of business scenario and liability scenario simulations before the optimizer is deployed. But, accounting within the model often is done while the optimizer seeks a feasible solution.

The sequence of model events runs like this:

1. Independently model many instances of exogenous states of the business world (e.g., asset returns, inflation, measures of economic activity, monetary conversion rates). Number these instances, $B_1, B_2, B_3, \ldots, B_n$. Note that each of these instances is a vector containing period-specific values for each operating fiscal period in the analysis.
2. Model many instances of the company's performance. Number these instances $C_1, C_2, \ldots, C_n$. $C_1$ often is dependent on $B_1$ because it may use an economic aggregate such as inflation or economic productivity to influence $C_1$'s business growth or loss and expense inflation. Each $C$ is a vector spanning the same fiscal periods as $B$.
3. Observe that in some DFA models neither $B$ nor $C$ is necessarily scaled to the actual volume of business. They are unit rates of change for underlying volumes that are yet to be applied.

4. Let the optimizer search mechanism posit a vector of weights that distribute the volume of assets at $t_0$, the inception point for a forecast period.

5. Apply the accounting mechanisms used by the DFA model to beginning assets and account for the unit activities expressed in $B$ and $C$.[20] Do this accounting for each vector pair $\{B_1, C_1\}, \{B_2, C_2\}, \ldots, \{B_n C_n\}$ over the range of its time span.[21]

6. Calculate the metric used for the goal and any constraints as of the end of the fiscal period, if it is a metric such as economic value or surplus. If it is a flow-based metric such as portfolio duration or discounted GAAP income, derive the metric for the holding period results. This calculation is done for each business/company scenario pair. There are $n$ results; collectively they constitute a simulated sample.[22]

7. Return the required metrics for the sample to the optimizer. If the optimizer is deployed for EF calculation, the goal will be a sample statistic for risk, such as variance, semi-variance, or chance-constrained percentile or range. The sample average for the distribution developed in step (6) for the metric will be used within the constraint set.

8. The optimizer will repeat steps (4)–(7) until it has obtained a feasible set.

The optimizer uses a sample. The optimizer results have sampling error. Steps (1) and (2) are experiments. Let there be 10

[20] At this stage, the derivation of taxes would occur. As noted by Rowland and Conde [1996], the determination of federal income taxes is convoluted by the combined effect of discount rates, changes in loss reserves, varying underwriting results, and tax carryforwards and carrybacks.

[21] Some models may achieve computational efficiencies when economic scenarios are paired with E($C$) instead of with direct pairing to $C_1, C_2, \ldots, C_n$. When this is done, however, the variance of the metric being optimized will be reduced, and the minimum variance portfolio is likely to be different.

[22] If enough pairs are used, the chance that the model will converge improves.

repetitions of this experiment. Application of steps (1)–(8) will result in 10 efficient frontiers, each derived from a different experimental sample. It is likely that they will have different characteristics.

In a DFA experiment there are many draws from the urn; each simulation is another draw. The modeler gets distributional information about the contents of the urn by the experimental grouping of all the simulations. When enough simulations within each experiment are run, convergence of the distribution of results can be achieved. Since it is unlikely for the output distribution to be known, or necessarily capable of being parameterized, no a priori estimate is available. Instead, an empirical measure of convergence must be used.

The allocation of company assets among competing investment alternatives using a single efficient frontier calculation (based on a single experimental result) may seem to be similar to betting on the allocation among balls of different colors within the urn based on a single sample from the urn containing them. One may, or may not, be lucky. But you improve your luck by increasing the number of simulations.

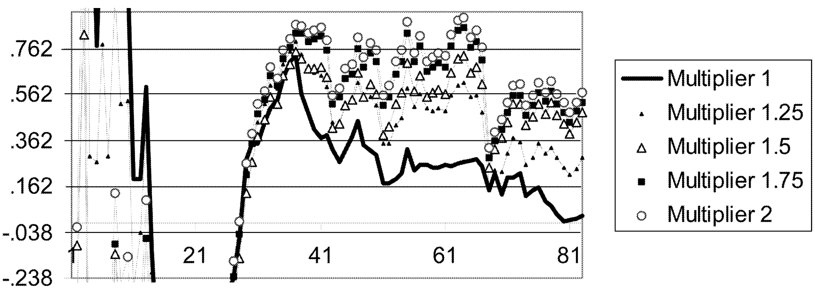
One still may become victimized by a faulty decision while ignoring sampling error. This may arise in calibrating a model to history. The historical record is a single draw from a true underlying probability distribution. We may be lucky that the number of periods in the historical realization contains sufficient information about the underlying process for unfettered decision making. But we could be victims of sampling error, which we are unable to control or even limit.

## 4. HISTORICAL PERFORMANCE COMPARISON

Figure 4 illustrates the performance of several portfolios over increasingly longer forecast periods. It shows results for portfolios, which, a priori, have different levels of risk for the same

FIGURE 4

COMPARISON OF PERFORMANCE FOR ON-FRONTIER AND OFF-FRONTIER PORTFOLIOS




**Performance Information Ratio**
**Historical Period: January, 1988 - December, 1992**
**Forecast: January, 1993 - December, 1999**
**Expected annualized return=.0825**

return.[23] The multipliers shown in the legend of Figure 4 are multiples of the minimum variance risk. The line for Multiplier 1 traces the performance of the on-frontier EF portfolio. Other lines in the figure with multipliers >1 show performance of portfolios with the same expected return but higher variance.

Figure 4 traces performance using a variation of the Sharpe performance measure.[24] It is known as the information ratio. The Sharpe performance ratio, which measures excess return to risk, is adjusted in the denominator of the information ratio. The denominator of the Sharpe performance indicator is changed to *excess* risk. The information ratio is given by:

$$\frac{E(r_p - r_f)}{SD(r_p - r_f)},$$

[23]Risk in this study is measured as the standard deviation of return.

[24]Laster [1998] created various portfolios by combining two asset components, domestic (represented by S&P 500) and foreign (represented by Morgan Stanley EAFE). His bootstrap samples of these two components were used to calculate portfolio variance,

where

$r_p$ = monthly return on the portfolio,

$r_f$ = monthly return on the risk free component of the portfolio,[25]

E = expectation operator, and

SD = standard deviation operator.

Although the information ratio was computed with monthly data, it is expressed as an annual measure in the paper.

### *4.1. EF Performance Is Better for Low Risk-Return Portfolios*

The off-frontier portfolios, so-called inefficient portfolios, achieve performance that rivals or betters that of the EF portfolio.[26] There is no concept of "significance" that can be attached to the observed differences. However, it is clear that the performance differences are great and that inefficient portfolios outperform the efficient one in the Figure 4. When performance is measured by geometric return, the underperformance of the EF portfolio can be more than 100 basis points, as shown in Figure 5. The underperformance shown in Figure 5 is measured over a seven-year holding period, and there was no portfolio rebalancing during this time. Data for other time periods and the use of intervening portfolio rebalancing might materially affect this evidence of underperformance.

The performance varies considerably with the level of return and historical period. For example, Figure 6 illustrates performance for an earlier period and a lower expected return level. Here, the EF portfolio does, indeed, outperform the off-frontier

---

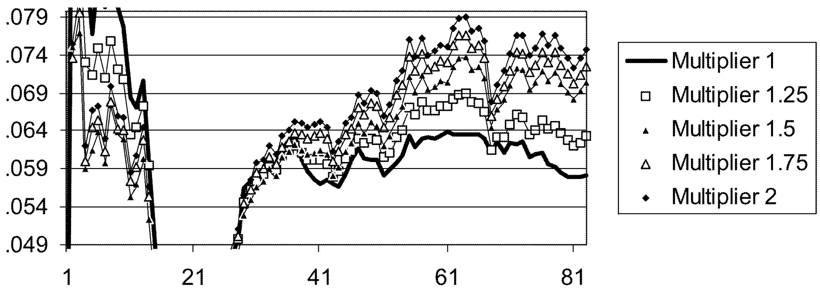
assuming various mixes. He did not separate historical and forecast periods. Instead, he measured quantiles from the bootstrap samples after constructing portfolios. He concluded that diversification into foreign equities substantially changed and improved the risk-return profiles.

[25]The 90-day Treasury bill index is used as the proxy for the risk free return.

[26]Short holding periods have performance measures calculated with few observations. The ordinal rankings among the different multipliers are volatile and should be ignored. The first six monthly periods are generally ignored in this paper.

FIGURE 5

COMPARISON OF GEOMETRIC RETURN FOR ON-FRONTIER AND OFF-FRONTIER PORTFOLIOS




**Geometric Return**
**Historical Period: January, 1988 - December, 1992**
**Forecast: January, 1993 - December, 1999**
**Expected annualized return=.0825**

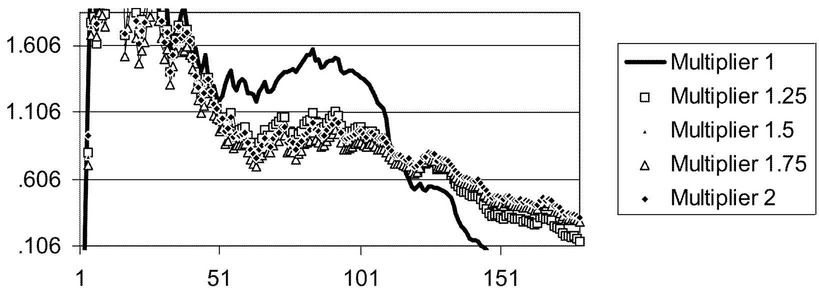
portfolios for about ten years. Thereafter, it reverses, and performance falls below off-frontier portfolios. The Figure illustrates that the contemplated holding period for use of an EF should probably not be as long. The performance variance illustrated in Figure 6 is volatile; the differences in performance in on- and off-frontier portfolios vary considerably with the choice of historical starting point and length of the holding period.

### *4.2. Overall Behavior of On-Frontier Portfolios for Information Ratio*

The historical record was examined from several perspectives to see whether an EF portfolio continues to outperform off-frontier portfolios. Equi-return portfolios were examined. These are portfolios whose returns are the same, but they have higher risk. The forecast period immediately following the end of the historical segment was examined to determine how long the on-frontier portfolio maintained superior performance. This forecast horizon extended to the end of the data, December 1999. Historical segments consist of a 5-year block of 60 observations.

FIGURE 6

EF PORTFOLIO PERFORMANCE AT LOW RISK-RETURN LEVELS




**Performance Information Ratio**
**Historical Period: January, 1980 - December, 1984**
**Forecast: January, 1985 - December, 1999**
**Expected annualized return=.0649**

Several adjustments were made for this analysis. The first six-month period was ignored because the ratio is highly volatile and computed from few observations. The extreme low return levels also were removed from the analysis because higher ones shown in the table dominated them.[27]

Table 1 shows the relative behavior of the information ratio at the return level indicated at the top of the column. Each row block includes the time for subsequent row blocks. For example, the forecast beginning January 1980 covers the period ending December 1999. The interval of measurement is a month. All of the other blocks begin at a later point, but all forecast periods end in December 1999.[28]

[27]The extreme low risk-return observations occur below where the EF curve has a positive first derivative. A portfolio with a higher return for the same risk can be found above this change in the curve.

[28]*Each block of rows uses a different set of on- and off-frontier portfolios—the respective EFs are derived from optimizations on different periods.* For example, the January 1980 forecast is based on the performance of EFs derived from a historical segment covering the 5-year period, January 1975–December 1979). However, the January 1995 forecast uses EFs derived from a different period, one covering the 5-year period, January 1989–

# TABLE 1

## INFORMATION RATIO BEHAVIOR

| Forecast Period | Return Levels | | | | | |
|---|---|---|---|---|---|---|
| Information Ratio (forecast begins 1/1980) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | 0.0095 | 0.0100 |
| Periods until on-frontier point underperforms (max = 238) | 6 | 6 | 6 | 6 | 6 | 6 |
| Number of periods on-frontier point outperforms all others | 10 | 142 | 148 | 153 | 154 | 151 |
| Average on-frontier rank (5 is highest) | 3.05 | 4.37 | 4.34 | 4.33 | 4.30 | 4.27 |
| Information Ratio (forecast begins 1/1985) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | 0.0095 | 0.0100 |
| Periods until on-frontier point underperforms (max = 178) | 111 | 110 | 109 | 7 | 119 | 69 |
| Number of periods on-frontier point outperforms all others | 105 | 104 | 103 | 8 | 113 | 124 |
| Average on-frontier rank (5 is highest) | 3.46 | 3.40 | 3.38 | 1.92 | 3.72 | 3.90 |
| Information Ratio (forecast begins 1/1990) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | 0.0095 | 0.0100 |
| Periods until on-frontier point underperforms (max = 118) | 40 | 6 | 9 | 9 | 9 | 9 |
| Number periods on-frontier point outperforms all others | 34 | 8 | 66 | 83 | 103 | 111 |
| Average on-frontier rank (5 is highest) | 4.18 | 4.05 | 4.57 | 4.72 | 4.89 | 4.96 |
| Information Ratio (forecast begins 1/1993) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | 0.0095 | 0.0100 |
| Periods until on-frontier point underperforms (max = 82) | 19 | 6 | 6 | 6 | 6 | 10 |
| Number of periods on-frontier point outperforms all others | 15 | 3 | 5 | 6 | 2 | 4 |
| Average on-frontier rank (5 is highest) | 2.10 | 1.83 | 1.82 | 1.81 | 1.60 | 1.56 |
| Information Ratio (forecast begins 1/1995) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | | |
| Periods until on-frontier point underperforms (max = 58) | 53 | 56 | 57 | Never | | |
| Number of periods on-frontier point outperforms all others | 47 | 50 | 51 | 53 | | |
| Average on-frontier rank (5 is highest) | 4.83 | 4.94 | 4.96 | 5.00 | | |

December 1994. The information in the blocks is not cumulative; the number of periods the on-frontier excels or outperforms off-frontier portfolios is a separate measurement for each row block. The row blocks show performance for portfolios constructed at different points in time.

Missing cells in Table 1 indicate that a feasible set was not found at that return level for one or more of the on- or off-frontier portfolios. There were five portfolios with risk up to two times the risk of the on-frontier point.

"Periods until on-frontier point underperforms" means the first period that an off-frontier portfolio beats the on-frontier efficient portfolio. "Number of periods on-frontier point outperforms all others" means the last period where the efficient portfolio wins. Performance tends to hold up better for lower return levels. This effect is reinforced by the larger values shown for the number of periods the on-frontier portfolio does outrank the off-frontier portfolios. In general, the on-frontier portfolio ranks well compared to the others. The average rank is generally high, above 3 out of 5. But the performance is not consistent. The on-frontier portfolio did well during the long forecast period starting January 1980 and during the shorter forecast period starting January 1995. However, the low average of the on-frontier for the January 1993 period shows that the performance is greatly influenced by the historical period and perhaps influenced by sampling error.

There also is great inconsistency in the number of periods before an off-frontier portfolio has a higher information ratio. The scan begins in period 6 of the forecast horizon, so the reversal shown in the table will either be never or a number between 6 and $n$. In most cases, the reversal is early, but not permanent. There are many situations where the on-frontier portfolio wavers between highest rank and something less. This latter fact is found in the rows, "Number of periods on-frontier point outperforms." In most cases this number is larger than the number of periods before reversion, indicating that the on-frontier waffles in and out of superior performance. This could be another indication of sampling error. The choice of an on-frontier point may not, and probably does not, imply superior performance.

### *4.3. Behavior for Other Performance Measures*

The information ratio is believed to be a valid measure of performance because it adjusts for variation in the return series during the period of measurement. Were it applied to two consultants' portfolio allocation recommendations, the consultant with *lower* excess returns could be ranked higher than the other consultant, because of proportionately *lower* risk in excess return. This may be small consolation to the holder of the lower wealth portfolio recommended by the higher ranked consultant. This is why it is important to assess other characteristics beyond the appetite for risk before making an allocation decision. The manager with the higher information ratio has the better cost of risk per unit of return; yet, it is not of much use if a minimum return level or ending wealth is required.

There is considerable historic instability in the standard deviation of returns. This can be seen in Figure 7, which shows the historic progression of changes in the standard deviation of monthly returns of the portfolio components used in this study. The lines show the change in standard deviation for rolling 5-year blocks of data.[29] Any performance measure that is a function of this risk proxy, such as the information index, will be inherently sensitive to such volatility and, perhaps, exhibit similar historic instability. This volatility in risk helps to explain why historical EFs may lack forecast power.

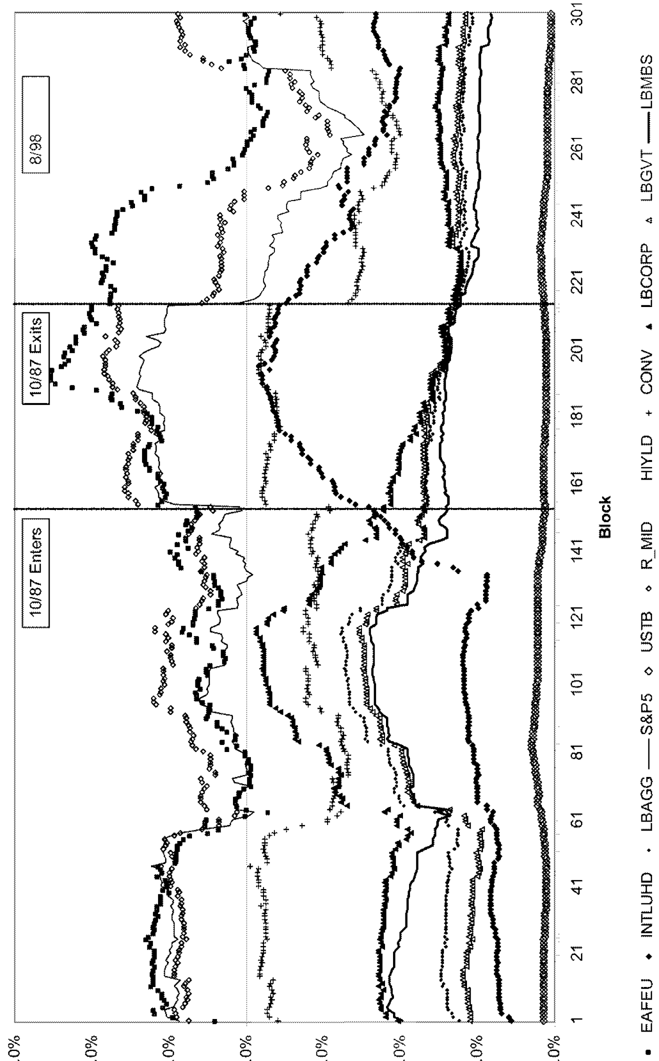
One measure of performance that is not risk-adjusted is geometric return during a holding period. Results are arrayed in Table 2. The layout of this table is similar to Table 1.

The forecast propensity of the on-frontier allocation is markedly changed. Wealth growth appears to be unrelated to the on- or off-frontier portfolio choice, and often is worse for the on-frontier allocation. The number of holding peri-

[29]There was significant volatility in the securities markets in 10/87 ("Black Monday") and 8/98 (Long Term Capital crisis). These periods are highlighted in the figure.

FIGURE 7

VOLATILITY IN RETURN STANDARD DEVIATION
ROLLING 5-YEAR MONTHLY STANDARD DEVIATIONS



## TABLE 2

### GEOMETRIC RETURN BEHAVIOR

| Forecast Period | Return Levels | | | | | |
|---|---|---|---|---|---|---|
| Geometric Return (forecast begins 1/1980) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | 0.0095 | 0.0100 |
| Periods until on-frontier point underperforms (max = 239) | 6 | 6 | 6 | 6 | 6 | 6 |
| Number of periods on-frontier point outperforms | 0 | 107 | 106 | 102 | 102 | 105 |
| Average on-frontier rank (5 is highest) | 2.41 | 3.35 | 3.39 | 3.38 | 3.38 | 3.40 |
| Geometric Return (forecast begins 1/1985) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | 0.0095 | 0.0100 |
| Periods until on-frontier point underperforms (max = 179) | 6 | 6 | 6 | 6 | 6 | 6 |
| Number of periods on-frontier point outperforms | 0 | 0 | 0 | 0 | 0 | 1 |
| Average on-frontier rank (5 is highest) | 1.00 | 1.00 | 1.00 | 1.00 | 1.03 | 1.04 |
| Geometric Return (forecast begins 1/1990) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | 0.0095 | 0.0100 |
| Periods until on-frontier point underperforms (max = 119) | 21 | 6 | 74 | 89 | 111 | 119 |
| Number of periods on-frontier point outperforms | 15 | 7 | 68 | 85 | 105 | 113 |
| Average on-frontier rank (5 is highest) | 4.11 | 4.06 | 4.60 | 4.75 | 4.92 | 4.99 |
| Geometric Return (forecast begins 1/1993) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | 0.0095 | 0.0100 |
| Periods until on-frontier point underperforms (max = 83) | 15 | 15 | 16 | 16 | 16 | 16 |
| Number of periods on-frontier point outperforms | 11 | 16 | 18 | 18 | 14 | 10 |
| Average on-frontier rank (5 is highest) | 1.99 | 2.19 | 2.22 | 2.22 | 1.92 | 1.73 |
| Geometric Return (forecast begins 1/1995) | 0.0066 | 0.0080 | 0.0085 | 0.0090 | | |
| Periods until on-frontier point underperforms (max = 59) | 6 | 6 | 6 | 6 | | |
| Number of periods on-frontier point outperforms | 0 | 2 | 9 | 30 | | |
| Average on-frontier rank (5 is highest) | 2.44 | 2.80 | 2.96 | 3.35 | | |

ods the efficient frontier portfolio dominates off-frontier portfolios is generally a lower proportion of the possible number of holding periods in Table 2 than in Table 1. Michaud [1998, pp. 27–29] claims there is a portfolio within the

EF, the "critical point," below which single period mean-variance efficient portfolios are also *n*-period geometric mean efficient and above which single period MV efficient portfolios are not *n*-period geometric mean efficient.

### *4.4. Performance Failure within CAPM*

Work with beta has led to various criticisms [Malkiel, 1996, p. 271].[30] For example, some low risk stocks earn higher returns than theory would predict. Other attacks on beta tend to mirror what we see with EF:

1. The capital asset pricing model (CAPM) predicts risk-free rates that do not measure up in practice.[31]
2. Beta is unstable, and its value changes over time.[32]
3. Estimated betas are unreliable.[33]
4. Betas differ according to the market proxy they are measured against.[34]
5. Average monthly return for low and high betas differs from predictions over a wide historical span.[35]

---

[30] Beta is a measure of systematic risk either for an individual security or for a portfolio. High beta portfolios, measured ex ante, in theory should have higher returns ex post than low beta portfolios.

[31] When ten groups of securities, ranging from high to low betas, were examined for the time period 1931–65, the theoretical risk free rate predicted by CAPM and actual risk free rates significantly diverged. Low risk stocks earned more and high risk stocks earned less than theory predicted [Malkiel, 1996, pp. 256–7].

[32] During short periods of time, risk and return may be negatively related. During 1957–65, securities with higher risk produced lower returns than low beta securities [Malkiel, 1996, pp. 258–60].

[33] The relationship between beta and return is essentially flat. Beta is not a good measure of the relationship between risk and return [Malkiel, 1996, pp. 267–8].

[34] Predictions based on CAPM about expected returns both for individual stocks and for portfolios differ depending on the chosen market proxy. In effect, the CAPM approach is not operational because the true market proxy is unknown [Malkiel, 1996, pp. 266–7].

[35] The ratio of price to book value and market capitalization did a better job of predicting the structure of nonfinancial corporate share returns than beta during a 40-year period [Fama and French, 1992].

Malkiel [1996, p. 270] concludes from his survey that, "One's conclusions about the capital-asset pricing model and the usefulness of beta as a measure of risk depend very much on how you measure beta." This appears to be true of EFs too. The *definition* of efficiency is what is important here—perhaps more important because correct measurement requires precise definition.

The choice of an optimization mechanism couched in terms of risk-return trade-off may not lead to wealth maximization. Under these pretenses one might wish to deploy a different optimization mechanism, such as the one mentioned by Mulvey, et al. [1999, p. 153] in which the optimization seeks to maximize utility. The choice of a particular utility function may be framed in terms of absolute risk aversion—negative exponential utility works in this regard.[36] And if the behavior of security prices does not have an observable stationary probability measure [Kane, 1999], utility approaches seem to be mandatory.

The subject of what is optimal is controversial and not apt to go away. The use of optimization within hybrid models and generation of metrics by DFA models has many subtle manifestations. One is the choice of planning horizon. Michaud [1998, p. 29] argues that investors with long-term investment objectives can avoid possible negative long-term consequences of mean-variance efficiency by limiting consideration to EF portfolios at or below some critical point. There is a parallel in our paper, in what we refer to as sampling error and its effect on the shape of the efficient surface. This surface appears to have properties at the lower risk-return areas of both lower dispersion, greater similarity in portfolio composition, and better on-frontier performance among different samples (either bootstrap or historic segment).

[36]The recommendation of a utility-decision approach has great breadth in the insurance literature—beyond the use of utility as goal function in optimization, other venues find it appropriate where stochastic dominance is sought. For example, exponential utility use was suggested in rate making by Freifelder [1976]. The choice of parameters for utility functions is perhaps as much an art as the parameterization of claims generations in DFA models.

## 5. CHARACTERISTICS OF THE EF SURFACE

The bootstrap-generated EF surface rises within the risk-return space. Views of this surface from two different angles are shown in Figure 8.

The surface is constructed from monthly returns. Looking down on the surface of the views, one obtains a projection on risk-return space. The surface is seen to curve as the efficient frontier curves. In the low risk-return sector, the surface is more peaked. The surface flattens and broadens in the risk-return space. Imagine yourself walking along the ridge starting in the southwest and proceeding northward and then northeast. You would first be descending a steep incline, and then a vista of a vast plane would unfold along your right. This can be interpreted within the context of changes in the marginal distributions representing slices through the surface either along the risk or along the return dimensions. We refer to the latter as an equi-return slice, and its properties are examined in more detail at a latter point in the paper. In either case, the visualization is one of moving from less dispersed marginal distributions to ones with greater variance as either dimension is increased.

There is an artifact of the intervalization that results in a sudden rise in the surface at the highest risk level. This occurs because higher risk observations were lumped into this final interval. Were higher levels of risk intervalized over a broader range, this ridge would flatten.

The surface shown in either of the views in Figure 8 is built from many efficient frontiers, each produced from optimizations done on a bootstrap sample. We already have seen in Figure 1 a subset of EFs that tangle together—they can be organized to produce a surface. The surface develops the same way an empirical probability distribution is built from a sample. Repeated sampling produces points that are intervalized and counted.

# FIGURE 8

## VIEWS OF EF SURFACE CREATED FROM BOOTSTRAP SAMPLES

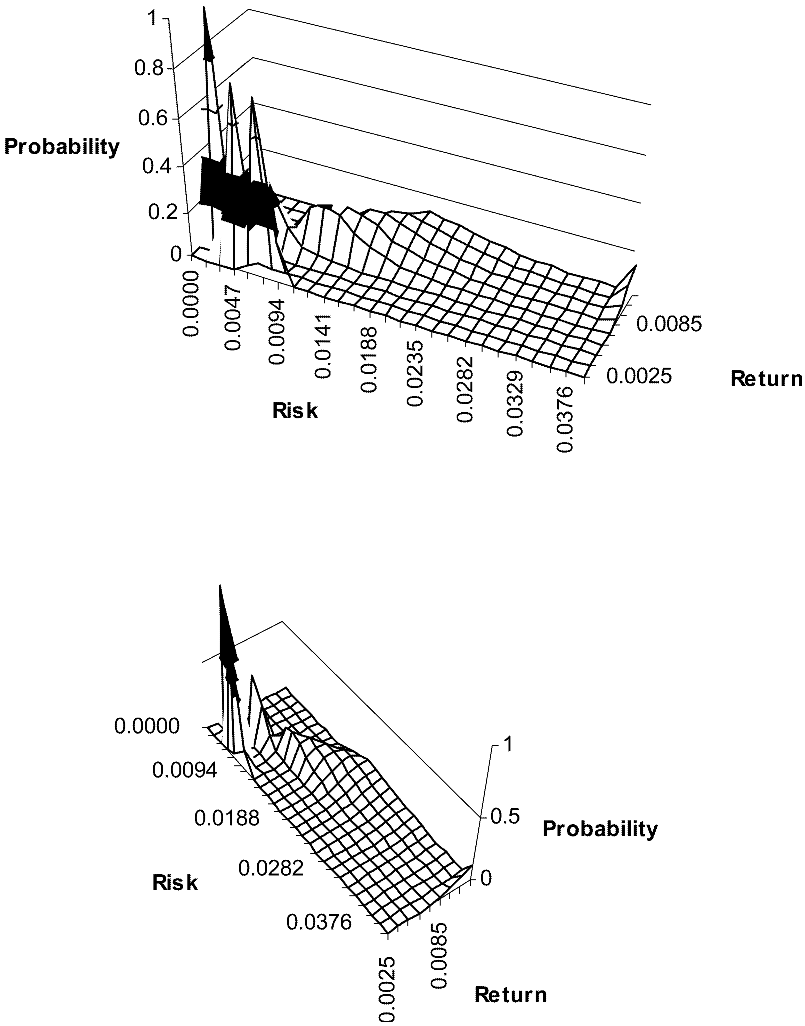
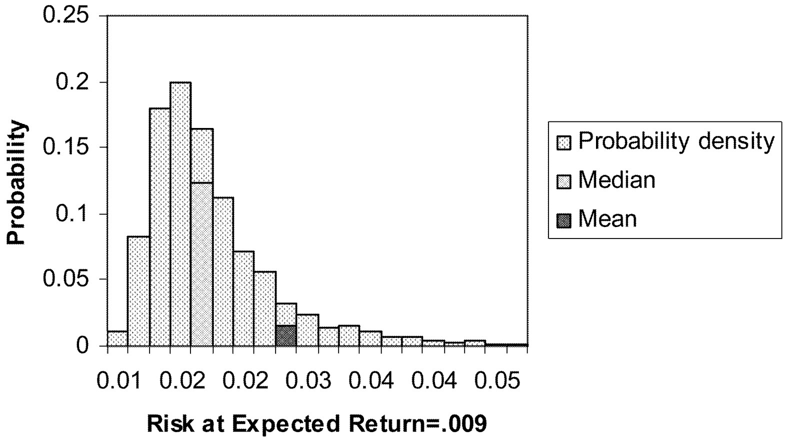


FIGURE 9

DISTRIBUTION OF RISK GIVEN A RETURN LEVEL




A frequency count can be made of observations for EFs falling within an arbitrarily small, two-dimensional region of risk-return space. An example of this mapping for 5,000 bootstrap-simulated EFs appears in Figure 8. Collectively, this mapping involves the two-dimensional intervalization of approximately 45,000 quadratic optimizations constituting the EFs for the underlying bootstrapped samples.[37]

### *5.1. Equi-Return Slice of the Efficient Surface*

A slice through the efficient surface along the return plane produces a histogram of the minimum risk points for a given return in the EFs used for the EF Surface. As return increases, this marginal probability distribution becomes more disperse. An example appears in Figure 9.

[37]Equi-return minimum variance points for the 5,000 bootstrapped EFs were intervalized based on an overall evaluation of the range of risk among all points on all EFs. If an efficient set could not be identified for a return level, the observation was ignored. The marginal probabilities (risk-return) were normalized to the number of viable observations for that risk level. The number of viable optimizations exceeded 4,500 at each return level.

## TABLE 3

### Statistics for Equi-Return Slices of the Efficient Surface

| Statistic | Efficient Surface from Bootstrapped Efficient Frontiers | | | | | | |
|---|---|---|---|---|---|---|---|
| Return Level | .0053 | .0066 | .0080 | .0085 | .0090 | .0095 | .0100 |
| Mean (times 1.0E4) | .0125 | .0533 | 3.76 | 8.86 | 17.9 | 33.6 | 50.7 |
| Standard Deviation (times 1.0E4) | .627 | 3.77 | 38.2 | 58.6 | 81.8 | 109.7 | 131.6 |
| Skewness (times 1.0E8) | .000123 | .378 | 57.8 | 136. | 275. | 516. | 779. |

## TABLE 4

### Statistics for Equi-Return Slices of the Surface Shown in Figure 10

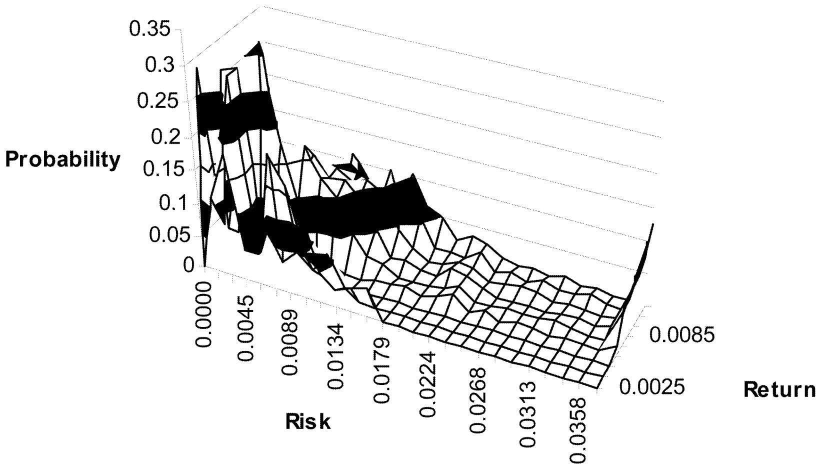
| Statistic | Efficient Surface from Historical Samples | | | | | | |
|---|---|---|---|---|---|---|---|
| Return Level | .0053 | .0066 | .0080 | .0085 | .0090 | .0095 | .0100 |
| Mean (times 1.0E4) | .663 | 5.53 | 21.0 | 25.5 | 28.8 | 33.2 | 46.5 |
| Standard Deviation (times 1.0E2) | .080 | .451 | .861 | .940 | .995 | 1.06 | 1.23 |
| Skewness (times 1.0E6) | .00676 | .769 | 2.92 | 3.54 | 4.00 | 4.62 | 6.46 |

The dispersion increases with return for both surfaces constructed from bootstrap samples and from randomly selected blocks of history. The distributions are positively skewed, increasingly so as return increases. The inset bars in Figure 9 identify the intervals containing the mean and median points of the distribution. Additional statistics both for bootstrapped and historical segment evaluations of sampling error appear in Tables 3 and 4.

The statistics are visually apparent in the EF surface shown in Figure 8. The surface is partially bowl-like—sloping downward in a concave fashion. Its rim encompasses a plane within the risk-return domain that is broad in the risk dimension. As one moves from low to high return, the marginal distribution of EF

FIGURE 10

EFFICIENT SURFACE FROM HISTORICAL SAMPLES

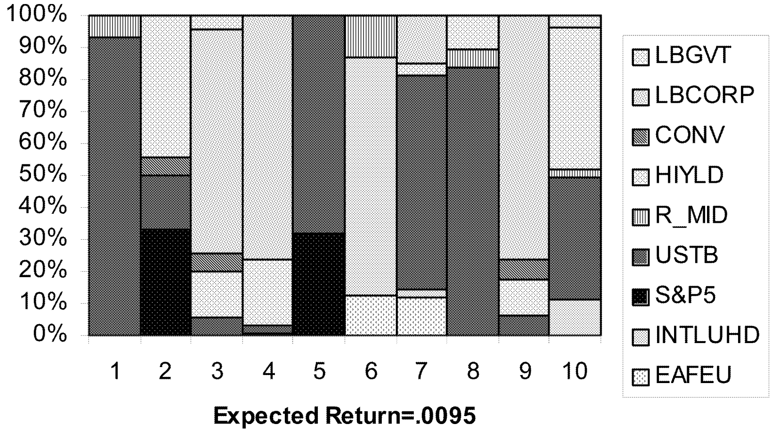



points measuring optimized risk (an equi-return slice through the surface as illustrated in Figure 9) becomes more dispersed. In a visual context as one moves from low to high risk along the EF surface and takes equi-return slices through it, one would find higher variance in the distribution of optimized EF risk points—variance shown in histogram plots such as Figure 9 is greater.

An efficient surface also can be created from EFs calculated for historical time periods. An example appears in Figure 10. The data are for 5-year overlapping blocks calculated on a monthly basis starting in 1970. The same general features are found in this representation of sample error. However, the surface is less flat than the one developed from bootstrap samples. The reduced dispersion in the surface of Figure 10 arises in part from the use of overlapping 5-year blocks used to construct the underlying EFs from which the surface is built. A statistical table similar to Table 3 was constructed for this surface. It appears in Table 4.

FIGURE 11

AVALANCHE CHART FOR HISTORICAL SEGMENTS




## 6. STABILITY OF PORTFOLIO COMPOSITION ALONG AN EFFICIENT FRONTIER

Portfolio allocation among component securities changes, usually dramatically, along the efficient frontier. A component may enter the feasible set at some point, increase in weight, decrease, and then drop out at other points along the EF. This effect was shown in Figure 2.

The change in composition for an equi-return level was examined among different EFs, constructed both from historical segment EFs and bootstrap EFs. We refer to this type of comparison as an avalanche chart because, when shown in an animation, the change in composition is similar to an avalanche. An example appears in Figure 11.

The vertical bars are stacked columns. Each segment within a column represents a different component of the portfolio. A br, therefore, compares the percentage value each component in the feasible set contributes across all components in the set. All bars are shown for a constant, equi-return level of an EF; but each bar is for a different historical segment. In Figure 11, each bar represents the portfolio composition for the equi-return level point on the EF, which was calculated for a 5-year block of monthly observations. The bars are for ten randomly chosen historical segments.[38] Were the blocks within the bars to consist of the same components and were they to be about the same size, the portfolio allocations would be the same regardless of the time frame. Examination of Figure 11 shows that the composition of the bars and individual component allocations varies considerably.

The portfolio composition is much more stable at lower risk-return levels. This result is in accordance with other similar findings based on the EF surface. It, too, shows less disperse results for lower return levels. This approach to measuring sampling error implies that performance of efficient frontiers may not be optimal relative to off-frontier portfolios. If the mix and composition of portfolios fluctuates considerably both with respect to historical and bootstrap sampling methods, the performance expectations of an ex ante allocation are not apt to hold ex post.

## 7. CONCLUSION

The behaviors shown in both Tables 1 and 2 illustrate a marked tendency towards randomness. The efficient surface built from bootstrap samples is highly variable within the risk-return domain. There appears to be some temporal dominance of on-frontier portfolios for lower risk-return levels, but the historical record is mixed. The bootstrapping of the single sample of asset returns provided by the historical data illustrates that sampling

[38]There is a small chance that two or more bars in an avalanche chart could be identical. However, there is a much larger probability that two or more bars have overlapping time periods in the calculation of their respective EFs.

error could materially affect the position and shape of the efficient frontier.

### *7.1. Should Efficient Frontiers Be Used in DFA Models?*

There is no strong support in this paper for the practical deployment of efficient frontiers in DFA. The risk in DFA models stems from model, process, and parameter risk. It affects all aspects of DFA models of the insurance enterprise. The existence of model and process risk [Kirschner and Scheel, 1998] thwarts the usual convergence to the true underlying distributions gained by running large numbers of simulations. When all of these new risk elements are heaped on top of the sampling error derived from asset model calibration or empirically measured covariance matrices, one wonders whether EFs are really useful in DFA analysis.

The work of Michaud [1998] bears on the issue of improving the performance of EF portfolios. He defines a measure of statistical equivalence for mean-variance efficiency. Any portfolio within the efficient surface *sufficiently close* to the optimal portfolio is considered equivalent to it. The extension of his idea to the efficient frontier surface is to identify a region on it whose ex ante chance-constrained probability both can be measured and has desirable statistical properties in a forecasting sense. This is analogous to acknowledging the existence of sampling error and specifying an unknown population parameter only to within an interval of statistical confidence. Unfortunately, the definition of *sufficiently close* is constructive but difficult to implement in a rigorous manner, particularly within the context of the hybrid DFA model.

Future study will have to answer the question of whether on-frontier asset allocations that are measured from hybrid DFA models suffer a similar unreliability. But the problems with on-frontier *asset* portfolios raised in this paper are apt to be exacerbated by inclusion of known sampling error in the liability side of DFA models.

### *7.2. How Can EFs Be Efficiently Deployed?*

Users of this construct should be aware that the term "efficient" in efficient frontiers has a good chance of being operationally false. The efficiency of portfolio composition is unlikely to be manifest in better performance of the on-frontier portfolio compared to other, off-frontier portfolios. The risk-return surface is not adequately measured by a single EF, and sampling error may lead to unwarranted conclusions about the efficacy of portfolios measured in such singular optimizations.

The user of EFs should probably view them as containing provisional, useful information about risk-return relationships. But, any single EF has limited value in understanding the risk-return surface. The conceptual basis of an efficient surface is an organized resampling of the data so that the decision process benefits from better understanding of uncertainty that might arise just because the EF is operationally derived from a sample. The misunderstanding of this uncertainty may lead to erroneous decisions, and the practitioner must be alert to potential inefficiencies of a single EF measurement. The authors recommend the elicitation of an efficient surface because the surface is apt to show a lack of statistical confidence in any single frontier on that surface. Under these circumstances, the practitioner must think in terms of confidence ranges. The sampling error shown in the efficient surface emphasizes how careful one must be when drawing inferences derived from optimization. An optimized frontier is based on an empirical covariance matrix, one that has sampling error. That error may be very important. It is easy to believe that strategic or tactical decisions motivated by so-called optimized DFA measurement will effectively move the organization to a better position in risk-return space. Unfortunately, there appears to be a broad region of "inefficiency" that may serve as well. An EF may be better than a crystal ball; but there is a good chance that it should not be taken too seriously.

## REFERENCES

[1] Chopra, Vijay, and William Ziemba, "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice," *Worldwide Asset and Liability Management*, Cambridge University Press, 1998, pp. 53–61.

[2] Correnti, Salvatore, Stephen M. Sonlin, and Daniel B. Isaac, "Applying a DFA Model to Improve Strategic Business Decisions," Casualty Actuarial Society *Forum*, Summer 1998, pp. 15–52.

[3] Davison, A. C., and D. V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge University Press, 1997, p. 582.

[4] Efron, Gerald F., and Robert J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hill, 1993, p. 436.

[5] Elton, Edwin J., and Martin J. Gruber, *Modern Portfolio Theory and Investment Analysis*, John Wiley & Sons, Inc, 1995, p. 715.

[6] Fama, Eugene F., and Kenneth R. French, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, June 1992.

[7] Freifelder, Leonard R., *A Decision Theoretic Approach to Insurance Ratemaking*, Irwin, 1976, p. 141.

[8] Kane, S.A., "What Happened to the Utility Functions? Imprecise Expectations of Security Prices," *International Review of Financial Analysis*, 8:2, 1999, pp. 165–175.

[9] Kirschner, Gerald S., "A Cost/Benefit Analysis of Alternative Investment Strategies Using Dynamic Financial Analysis Tools," Casualty Actuarial Society *Forum*, Summer 2000, pp. 21–54.

[10] Kirschner, Gerald S., and William C. Scheel, "The Mechanics of a Stochastic Corporate Financial Model," *PCAS* LXXXV, 1998, pp. 404–454.

[11] Laster, David S., "Measuring Gains from International Equity Diversification: The Bootstrap Approach," *The Journal of Investing*, Fall 1998, pp. 52–60.

[12] Levy, Haim, and Deborah Gunthorpe, "Optimal Investment Proportions in Senior Securities and Equities Under Alternative Holding Periods," *The Journal of Portfolio Management*, Summer 1993, p. 31.

[13] Malkiel, Burton G., *A Random Walk Down Wall Street*, W. W. Norton & Company, Inc., 1996, pp. 227–276.

[14] Michaud, Richard O., *Efficient Asset Management*, Harvard Business School Press, 1998.

[15] Michaud, Richard O., "The Markowitz Optimization Enigma: is 'Optimized' Optimal," *Financial Analysts Journal*, 45:1, pp. 32–42.

[16] Mulvey, John M., Chris K. Madsen, and François Morin, "Linking Strategic and Tactical Planning Systems for Dynamic Financial Analysis," Casualty Actuarial Society *Forum*, Summer 1998, pp. 149–167.

[17] Rowland, Vincent T., and Frank S. Conde, "Dynamic Financial Analysis Issues in Investment Portfolio Management," Casualty Actuarial Society *Forum*, Spring 1996, pp. 205–240.

[18] Sharpe, William F., "The Sharpe Ratio," *The Journal of Portfolio Management*, Fall 1994, pp. 49–58.

## APPENDIX A

## Review of Data Sources

This paper uses monthly time series of asset class total returns. A selection of broad asset classes typical of P&C insurance company asset portfolios was chosen for examination. The time series all begin January 1, 1970. However, certain asset classes (e.g., mortgage backed securities) do not have a history that extends back this far. For these classes, the time series were backfilled to the January 1, 1970 start date by an investment consultant. The backfill process was based on a consideration of the market conditions of the time (e.g., interest rates, fixed income spreads, inflation expectations) and how the particular sector would have performed given those market conditions. The Start Date in Table 5 refers to the date historical data begin.

### TABLE 5

### Asset Components

| Class | Code | Source | Start Date |
|---|---|---|---|
| International Equities | EAFEU | MSCI EAFE Index | 1/1970 |
| International Fixed Income | INTLHDG | JP Morgan Non-US Traded Index | 1/1970 |
| Large Cap Domestic Equities | S&P5 | S&P 500 Index | 1/1970 |
| Cash | USTB | 90 Day US Treasury Bill | 1/1970 |
| Mid Cap Domestic Equities | RMID | S&P Mid Cap 400 Index | 1/1982 |
| High Yield | HIYLD | CSFB High Yield Bond Index | 1/1986 |
| Convertible Securities | CONV | CSFB Convertible Index | 1/1982 |
| Corporate Bonds | LBCORP | Lehman Brothers Corporate Bond Index | 1/1973 |
| Government Bonds | LBGOVT | Lehman Brothers Government Bond Index | 1/1973 |
| Mortgage Backed Securities | LBMBS | Lehman Brothers Mortgage Backed Securities Index | 1/1986 |

## APPENDIX B

## Annualized Returns

The time series used in this study are monthly returns. With the exception of work relating to performance, all returns are expressed as monthly returns.

For performance measurement purposes, returns have been annualized using the following formulas.

*Annualized Expected Return*

$$R_p = (1 + r_p)^{12} - 1,$$

where

$R_p$ = annualized return, and

$r_p$ = monthly return.

*Annualized Variance of Return*

$$V_p = [v_p + (1 + \mu_p)^2]^{12} - (1 + M_p)^2,$$

where

$V_p$ = annualized variance of return,

$v_p$ = monthly variance of return,

$\mu_p$ = expected monthly return, and

$M_p$ = expected annualized return.

*Annualized Geometric Return*

The growth rate, $g$, for a holding period of $n$ years is given by:

$$1 + g = \left(\frac{V_n}{V_0}\right)^{1/n},$$

where

$V_n$ = portfolio value at the end of the holding period $n$, and

$V_0$ = portfolio value at the beginning of holding period.