Asset Allocation and Long-Term Returns: An Empirical Approach

Stephen Coggeshall and Guowei Wu
Morgan Stanley
July, 2005

Executive Summary

We present a discussion of the standard approaches to asset allocation, generally falling into two camps: Mean Variance Optimization (MVO) and the maximization of the final value of a wealth utility function (UF). After describing shortcomings in both of these standard approaches, we describe a heuristic, empirical approach that uses concepts of Shortfall Risk as an objective and actual data as a direct model of the stochastic market evolution. This empirical approach leads to fundamentally different conclusions than the standard approaches, and by definition fits the past data much better.

Section 1 describes the mainline approaches to asset allocation, MVO and UF, with mathematical details in Appendix A. This appendix provides concise summaries of many special cases of both MVO and UF. In Section 2 we begin the description of the empirical approach, first discussing the validity of using overlapping windows and a variety of ways practitioners measure returns and standard deviations, ending with a recommended methodology to measure these statistical properties for both short- and long-term time horizons. Section 3 contains discussion on the underlying nature of statistical market dynamics, in particular their deviation from a pure random walk.

Section 4 shows the results of applying the data-driven analysis to long-term stock holdings and quantifies the deviation from random walk for long-term returns and its implications. In Section 5 we have a brief discussion of risk and alternative ways of thinking about risk. Section 6 describes the long-term nature of the bond market and compares it to equities. Section 7 describes a new, empirical approach to asset allocation using Shortfall Risk as a performance goal. We discuss the philosophy behind using empirical verses theoretical approaches in Section 8.

We find that

- For holding periods of about 15 years and greater, bonds are riskier than stocks in the practical sense that bonds are always expected to under perform stocks over 15-year or higher time horizons,
- While short-term stock returns show fat tails, long-term stock returns show *skinny* tails,
- While stocks show mean reversion, bonds exhibit mean *aversion*.

1. Standard Approaches to Asset Allocation

The standard approaches to asset allocation involve a combination of data and theory. One examines stock price evolution over some historical time window, calculates several key statistical properties (return and standard deviation of return), and then uses theory (lognormal prices, random walk) to project to long-term market performance. There are then two primary families of approaches to identify optimal asst allocation: (1) Mean Variance Optimization (MVO) and (2) Optimization of a utility function. In both of these approaches one uses the two statistical measures calculated from the historical window (mean, standard deviation or sigma), and assumes a random walk model with lognormal prices to both scale the two statistical measures to an annualized basis and to extend to long-term evolution.

1.1 Mean Variance Optimization

The general concept in MVO is to use the two statistical measures (mean, sigma) along with lognormal price random walk theory, in a linear approximation of asset relationships (correlation coefficients from the covariance matrix) to construct a curve in risk/return space called the Efficient Horizon that identifies the optimal combinations of assets. This is typically done for a one-year time horizon, but one can extend the methodology to any time horizon, as well as allow multiperiod reallocation. The optimal portfolio choice (asset allocation) will be some location on this Efficient Horizon, chosen through some quantification of risk tolerance. Typically, this risk tolerance quantification in the MVO approach is ad hoc.

1.2 Utility Function Approach

A different approach to identify optimal asset allocation is through the use of a utility function, which quantifies the value of wealth to an individual. One writes down an analytic objective function to optimize the utility of wealth over all possible choices of asset allocation, in a stochastic return environment. This methodology can readily include a variety of important effects, such as time-dependent penalties for not achieving particular results. To calculate optimal asset allocations one needs to specify all evolutionary processes, penalties and the form of the wealth utility function.

Appendix A contains details and results from these two standard approaches to asset allocation.

1.3 Results and Issues with These Two Standard Approaches

Both of these traditional approaches have advantages and drawbacks. The MVO approach relies completely on two statistical measures, the mean return and the standard deviation (sigma) of the different asset classes. We can think of sigma (or the variance, sigma squared) as a matrix of correlations, where the diagonal is the single asset variance. The reliance on these linear statistical measures means that MVO only considers linear relationships between asset classes, and the market dynamics are assumed to be random walk with a lognormal distribution. The objective in MVO is to maximize return for a given risk, or alternatively, to minimize risk for a given return. The set of constraints that can be handled by this method are limited, and it is difficult to add more realistic objectives such as consumption or penalties for not achieving certain goals. The output of MVO is a curve called the Efficient Frontier that spells out the family of returns for given risks. One then needs to select a point on this curve to operate, and the selection of this point is related to one's risk tolerance. The risk tolerance along the Efficient Horizon is not quantitatively identified and is only a qualitative choice.

MVO is usually done in a single period, usually one year. However, it is possible to do a multiperiod MVO for an arbitrary time horizon, and the result is a time-dependent, path-dependent asset allocation strategy. Multiperiod MVO gives a rule for how to reallocate over each period depending on the results of the previous periods. Obviously, MVO is very sensitive to the values of the mean and sigma.

The utility function (UF) approach allows the use of much more complex objective functions, in fact, arbitrarily complex. One can proceed analytically with very complex objective functions, which frequently include time-dependent consumption and financial needs, such as college costs or a house. The UF optimization results in an asset allocation over time and also produces the optimal consumption through time. There are a few standard UFs, generally having a fixed risk parameter gamma.

There is disagreement on the fundamental assumption in the UF approach, the assumption that there exists a global UF that begins at the origin and moves out smoothly, concavely down. Behavioral finance experiments by Kahneman and Tversky (2000) have shown that real UFs are much more asymmetric and locally more concave downward than any global function, and disprove the assumption that a global function exists. Further, the risk aversion is quantified by the concavity, which is a local property and cannot be represented analytically globally. Similar conclusions have been discussed by Rabin and Thaler (2001). See also Stutzer (2003) for a good discussion on the competing approaches of Shortfall Risk and utility function, as well as a possible methodology to bridge the gap by allowing Shortfall Risk to be a goal in a utility function approach by allowing the risk parameter gamma to also vary in time.

The UF approach can naturally be done in a multiperiod framework. However, to build the necessary recursion relationship to solve the problem analytically one must assume that the risk aversion gamma is constant over time. As in the MVO approach, there is no explicit relationship of the parameter gamma to a quantitative measure of risk. Consequences of the UF approach under the most common CRRA UF choice are that the fraction of cash to risky assets is determined by gamma, and then the allocation between risky assets is determined by the sigmas. All these allocation fractions, in simple and complex cases, remain constant over time regardless of the choice of gamma, which is not realistic. Regardless of one's risk aversion, one generally and properly moves from risky to less risky assets as one's time horizon shrinks. The UF approach does not have this as a consequence, and so the relevance of this approach for practical use is questionable.

2. Alternative Approach: Directly Use Data

The standard approaches make use of a combination of historical data and theory, which of course is guided by data. Inherent in these approaches is an assumption that the future will behave as it did in the past, albeit stochastically with parameters obtained from the past data. There is a heavy reliance on the selected theory to generate the relevant statistical parameters, but more importantly to extrapolate to long-term evolution.

An alternative approach is to use the data directly to obtain actual historical distributions and use these empirical distributions to stochastically simulate performance, without the use of a theoretical structure. The benefit of this data-driven empirical approach is that one doesn't need to construct a theoretical framework, either simple to allow analytic treatment or more complex requiring numerical solutions, and therefore rely on the consequences of the theoretical assumptions. A drawback of a data-driven approach is the need to justify that the past is representative of the future. We submit that this need is also just as present in the standard approaches, because not only are the calculations of the statistical parameters (return, sigma) completely dependent on history, but just as importantly the choice of the theoretical construction is justified by examining historical evolution. Further, we submit that no theoretical framework can reproduce/match the past better than using the past directly, as long as it can be done in a statistically meaningful manner.

2.1 Overlapping Windows are OK

A common concept in the analysis of financial data is that one should not use overlapping windows in calculating statistical properties, such as returns. For example, if one has several years of daily time series of a stock price, and one would like to calculate important statistical properties (annual return, sigma, distribution of return), a common approach would be to use monthly values at some fixed day of the month, say the first or last day of the month. A common method is to use this series of monthly values to calculate an average return by averaging these monthly returns and then multiplying the result by 12 to annualize. This would be a nonoverlapping window approach, since the time windows are completely nonoverlapping; each one uses a separate time period with no overlap (see Figure 1).

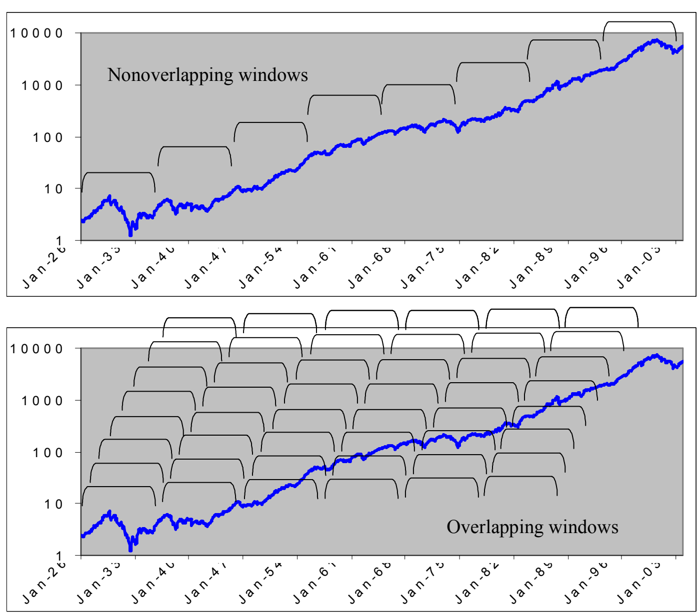


Figure 1. The difference between nonoverlapping and overlapping windows.

We assert that using overlapping windows is not only statistically valid, but that it does nothing but improve the statistical measures, even though the additional information is indeed related to the nonoverlapping data. **In any instance where using nonoverlapping windows is statistically valid, using similar but overlapping windows only improves the calculation.**

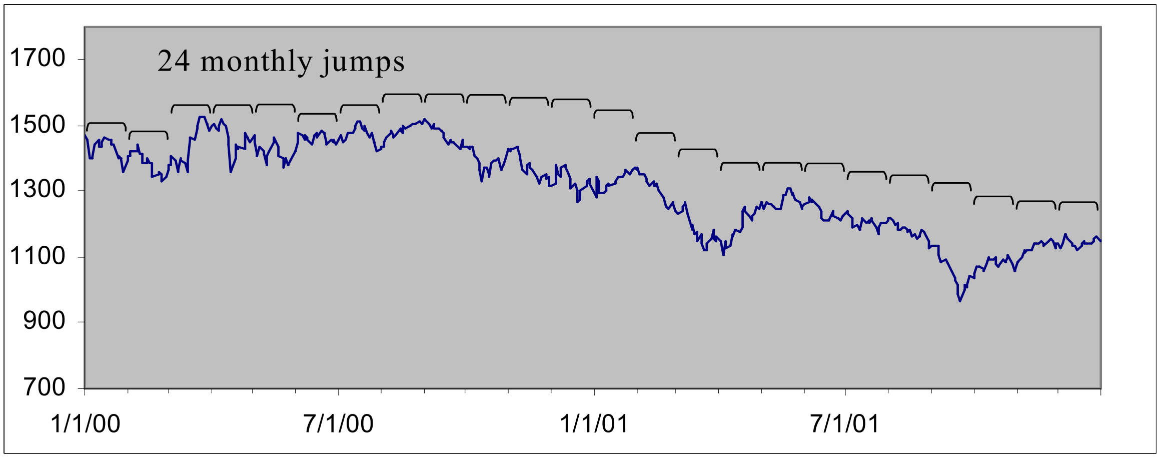


Figure 2. Two-year time series of S&P price with 24 nonoverlapping windows.

Consider a 2-year time series of S&P data, shown in Figure 2, along with the 24 nonoverlapping windows. It is not uncommon for one to calculate an average annual return for such a series by averaging these 24 monthly returns and multiplying by 12. If one wants to estimate the distribution of returns, one has 24 discrete data points in which to estimate this distribution. When presented with limited data, one always smooths the data in some way, as shown in Figure 3. Perhaps the most common smoothing technique is to simply bin the data, shown in the upper right of Figure 3. This is essentially what one does when one uses the mathematical formula for the mean and standard deviation using these 24 data points.

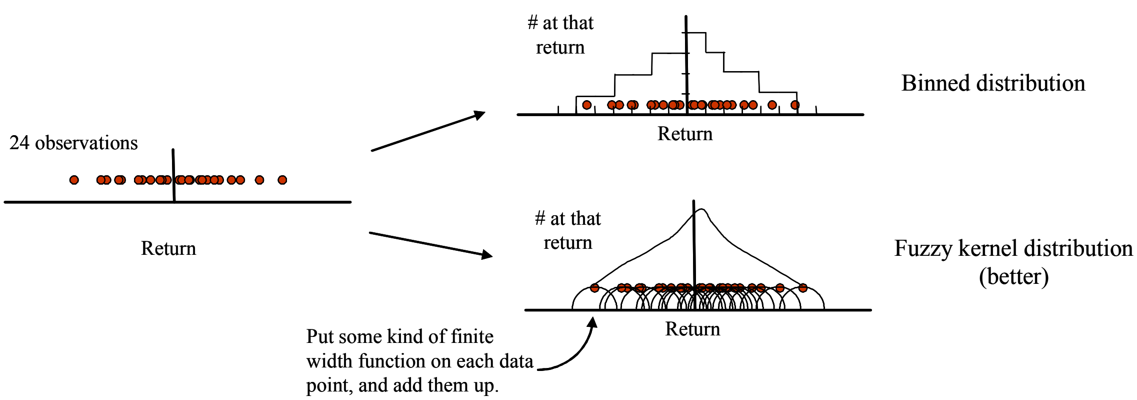


Figure 3. When presented with a finite data set, one always uses some form of smoothing.

A generally better way to smooth finite data is to use a kernel function smoothing technique. Here one decides some sort of finite-width function shape and places one on each data point. We then just add up all of these kernel functions and get a smooth distribution. The exact shape of the kernel function doesn't matter much; popular ones are tent functions (simple triangles), Gaussian (normal distributions), or semicircles (shown in Figure 3). The choice of width is important: too wide and the data is smoothed too much; too narrow and there is no smoothing. A good choice is a few times the average distance between points. Note that binning is simply using a flat kernel function, but in binning the kernel function is not centered on the data point but tied to the grid.

Another way to smooth the data points in finite data is to "jitter" them a little to the right and left. One can do this using a variety of ways to choose this shifting distance right or left. Essentially, this is what the kernel functions are doing, and the kernel function shape represents the probability distribution function of the "jittering" distance. With this in mind, one can choose a more realistic way to do this "jittering". If we shift the monthly window by a day or two in either direction of the original position on the data, we have "jittered" the return data point by a realistic value without presupposing any kernel function shape. In fact, we can shift this window the maximum amount in either direction (a half month), which provides optimal natural smoothing. This is identical to using a shifting overlapping set of windows.

Next we examine the mathematics of using such shifting, overlapping windows as opposed to nonoverlapping windows. In the above case we have 24 return data points R1, R2… R24. From this list we wish to compute the relevant statistical measures such as mean return, standard deviation, return distribution, etc. Consider the case where we simply duplicate each of the 24 return values 24 times each (25 days in a month[*]), so we now have 24 x 25 data points, but many of them are identical. This process is shown in Figure 4.

Start with list of 24 nonoverlapping returns R1 R2 R3 R4 R5 R6 R7 R8 R9 …	Consider a similar list of returns where each entry is duplicated 24 times (25 days in a month) R1, R1, R1, R1, … R2, R2, R2, R2, … R3, R3, R3, R3, … R4, R4, R4, R4, … R5, R5, R5, R5, … R6, R6, R6, R6, … R7, R7, R7, R7, … R8, R8, R8, R8, … R9, R9, R9, R9, … …	Now, modify this second list by replacing the duplicated returns with returns shifted by one day (overlapping windows) R11, R12, R13, R14, … R21, R22, R23, R24, … R31, R32, R33, R34, … R41, R42, R43, R44, … R51, R52, R53, R54, … R61, R62, R63, R64, … R71, R72, R73, R74, … R81, R82, R83, R84, … R91, R92, R93, R94, … …
From this list compute • Distribution • Mean return • Standard deviation • Any statistical measure	From this list compute • Distribution • Mean return • Standard deviation • Any statistical measure This set of statistics is identical to the first one. Duplicating events uniformly doesn't change anything.	This set of returns is, by all measures, more accurate and more relevant than both the set of simply repeating numbers and the first set (nonoverlapping). This set of statistics is better than the first one. Adding additional relevant samples is better than a smaller set of samples.

Figure 4. The mathematics shows that using overlapping windows is better than nonoverlapping windows.

We first recognize that in using this expanded list of duplicated numbers, neither the mean nor standard deviation will be different from the original data set. Now we see that we can improve from this expanded list by modifying many of these new numbers to more accurate, realistic numbers using the data.

Finally, we recognize that the nonoverlapping window approach has the potential to miss important events, such as 9/11. In Figure 5 we show the S&P for the two year period from 1/2000 through 1/2002, with the monthly windows overlaid. The bottom left picture is a close-up of the plot around the 9/11 event, and we see that nonoverlapping windows starting on month beginning or end will miss most of this dramatic event in the market. This process will miscalculate the return average and distribution, and in particular will underestimate the standard deviation. A sliding, overlapping window approach will catch these relevant high frequency events that may be missed with nonoverlapping windows.

[*] The choice of this second 24 is because with 25 days per month, we can have 25 different starting points per month, so we add 24 new possible returns to get to 25 per month. This new 24 has nothing to do with the original 24, from 2 years monthly returns.

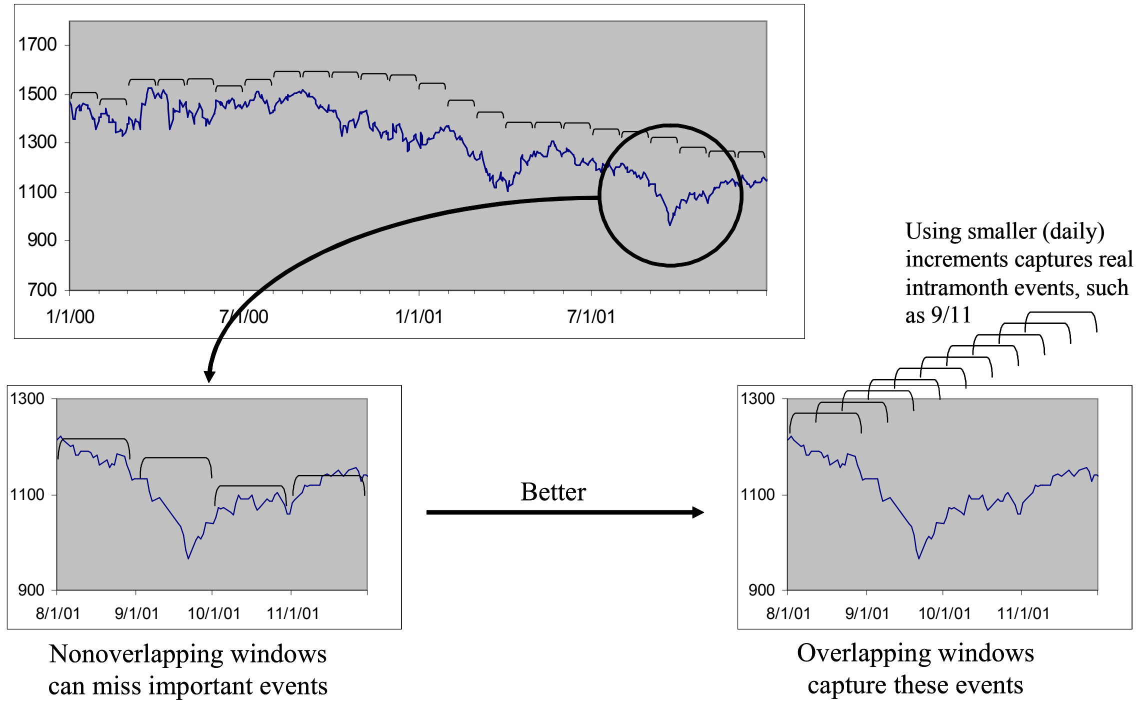


Figure 5. Nonoverlapping windows can underestimate/miss important events such as 9/11. These events are correctly captured with overlapping windows.

In conclusion, we submit that in any instance where one is comfortable using a nonoverlapping window approach, the results are only improved if one uses an overlapping approach. What we are suggesting is that adding higher frequency data only helps the accuracy of the statistics and never hurts. Why then not carry this to an extreme and use tick data? The results of our argument show that this will only help, but at some point the difficulty of the calculation outweighs any added benefit. We suggest using the highest frequency data that is readily available, but that for examination of long-term returns, going to intraday data will not affect the statistics in any meaningful way. However, for investigations of short-term dynamics, intraday data is very useful.

The usual argument for not using overlapping windows is that returns using a sliding, overlapping window are not independent. This is completely true, but as we have shown above, adding in this nonindependent information in all cases only helps and doesn't hurt; it only makes the statistical calculations more relevant, accurate and robust. On the other hand, we must keep in mind that even though using overlapping windows only helps, one should still be cautious about drawing conclusions about distributions that are formed from only a few examples. For example, if one has three years of monthly data, can one make any reasonable statement about an average yearly return? Probably, and many people do. How about if one only has one and a half years data? This is probably not enough to make reasonable statements about an average yearly return. More data is always better, but one needs a heuristic feel for when one can make reasonable statements. Perhaps a reasonable working rule of thumb might be, one needs on the order of a few nonoverlapping periods, then use overlapping to smooth it out. Even then we should be careful about over interpreting the distribution.

2.2 How to Calculate Mean Return and Standard Deviation of Return

We continue the discussion on using data nonparametrically by examining how one traditionally calculates the statistical properties return and sigma. Given a stock time series, how does one go about calculating the average return and the standard deviation (sigma) of this return?

We find that everyone agrees on the meaning of these numbers, average return and standard deviation, but there is much disparity in how they are calculated. Returns and sigmas are generally reported on an annualized basis, regardless of how they are calculated. Sometimes people report monthly returns, but this is clearly stated when it's done. We focus here on the annual return and sigma numbers.

There are many methodologies in how these two important numbers are calculated given a time series of a stock. One could put these into two general classes: (1) look at returns using time window sizes smaller than a year (e.g., a month or a day), and then use some process to annualize, or (2) use an annual window size, but generally use nonoverlapping windows. We look in more detail at these two approaches.

(1) Calculate annual returns using a smaller time window

The window sizes used are typically one month, but sometimes people look at daily returns and project to annual. For illustration purposes, consider the number of days in a year to be 252 and the average number of trading days in a month at 21. In this method, the window size N is smaller than 252, and is typically 21. A common method is to look at stock values at a fixed time each month, say the last day of the month. The return for that window is calculated through

$$R_i = \frac{S_i - S_{i-N}}{S_{i-N}}$$

where S_{i-N} is measured N days before S_i, e.g., 21 days. The next step is how to annualize this number. A common way is to simply calculate the arithmetic average, and then multiply this average by 252/N to get the approximate 252-day (annual) return:

$$R_{annualized} = \frac{252}{N} \times \frac{1}{L} \sum_{i=1}^{L} R_i,$$

in the case with returns over L different time windows.

A better way is to use geometric compounding of an arithmetic average:

$$R_{aa} = R_{arithmetic\_average} = \frac{1}{L} \sum_{i=1}^{L} R_i \quad ,$$

$$R_{annualized} = \prod_{j=1}^{252/n} (R_{aa} + 1) - 1 = (R_{aa} + 1)^{252/n} - 1,$$

which is the geometric compounding formula, and is more accurate. Both of these methods require the assumption that the stream of returns is iid, identical and independent, which is generally not true (e.g., trends can exist).

(2) Calculate annual returns using an annual window

In this method the window size N is equal to 252, so one is directly calculating an annual return, which we believe to be a more accurate method rather than using a smaller window size and using assumptions to project to annual. However, usually people mandate the use of nonoverlapping windows, so they have at most M values of this return with M years of data. Even with yearly data many people use the arithmetic average above as opposed to a geometric average, which can produce very wrong results. Consider the case where a portfolio loses 99% in the first year, then has a 100% gain the second year (doubles). A \$100 investment would be worth \$1 after the first year, and then \$2 after the second year. The arithmetic average annualized gain is 0.5%, which is very misleading compared to the more realistic geometric annualized return of -86%.

While we believe that this annual window calculation is substantially better than the smaller window size, many people require the use of nonoverlapping windows. Based on the arguments in the previous section we believe that it is always better to use overlapping windows. Our preference is to use this second method using a moving window size of one year to calculate the annual return directly, and slide the window by the shortest relevant frequency in the data. With this procedure, one can build the most accurate distribution of annual returns and also directly calculate the sigma of this distribution.

Next we consider the correct way to calculate the standard deviation (sigma) of this stream of annual returns. First, we recognize that it is wrong to simply calculate the sigma of the stream of annualized returns from Method (1) above. If one is using Method (1), what is typically done is to calculate a sigma of the N-window returns (σ_{N_Window}, N < 252), and then annualize this sigma by assuming a random walk. The most common way to do this is to assume the square-root-time scaling for such a process, and scale by that factor:

$$\sigma_{annualized} = \sqrt{\frac{252}{N}}\sigma_{N\_Window}.$$

This is actually an approximation to the exact formula under the lognormal random walk assumption:

$$\sigma_{annualized} = \sqrt{\left[\sigma^2_{N\_Window} + \left(1 + R_{N\_Window}\right)^2\right]^{252/N} - \left(1 + R_{N\_Window}\right)^{504/N}},$$

where R_{N_Window} is the average of the returns over the many different N-sized time windows, and σ_{N_Window} is the standard deviation of these returns over the N-sized windows. If we use Method (2) where N=252, we don't need to annualize, and so we can directly calculate the standard deviation of these annual returns exactly.

Table 1 below summaries these different methods. Our preference is to calculate both the annual return and sigma directly, as noted in the table.

Statistical Measure	How to Calculate	Problems
$\mathbf{R_{annualized}}$	Monthly average x 12	Arithmetic, but popular; requires iid assumption
$\mathbf{R_{annualized}}$	(arithmetic monthly average + 1)^{12} -1	Geometric (better); still requires iid assumption
$\boldsymbol{\sigma_{annualized}}$	σ of annualized returns, returns annualized as described in the first two table entries	Wrong[†]
$\boldsymbol{\sigma_{annualized}}$	$\sqrt{12}\sigma_{monthly}$	Approximate and popular; good for small σ and R; still requires iid assumption.
$\boldsymbol{\sigma_{annualized}}$	$\sqrt{\left[\sigma^2_{monthly} + (1 + R_{monthly})^2\right]^{12} - (1 + R_{monthly})^{24}}$	Better; still requires iid assumption
$\mathbf{R_{annualized}}$	Directly using annual skips, sliding window daily	Best
$\boldsymbol{\sigma_{annual}}$	Sigma of the above direct annual returns	Best

Table 1. Different methods to calculate annual returns and standard deviations.

Our preference is always to calculate the return and sigma for any N-day window directly without having to use an iid assumption. If we are interested in a 10-year return, we prefer to use sliding, overlapping windows of size 10 years.

[†] When projecting from a short-term measurement to a long-term-process statistical measurement we are **projecting** a process, and we need to use the correct formula
$\sigma_{annualized} = \sqrt{\left[\sigma^2_{N\_Window} + \left(1 + R_{N\_Window}\right)^2\right]^{252/N} - \left(1 + R_{N\_Window}\right)^{504/N}}$. Later in this paper we will **scale** a longer process to an annual scale, and we will do exactly as described in this table: we first scale all the returns to their equivalent annual scale, then calculate the sigma of this annual scale. It's important to note that scaling to a smaller timescale is very different from projecting a short-term process to an equivalent long-term process.

We should point out here a common discrepancy in usage. All the sigmas we have been discussing so far are for a finite time windows, and we project them to a one year time window when we annualize, so they are easier to compare. These finite-time window sigmas are different from the continuous-time, continuous-compounding returns and sigma used in the generation of continuous time frameworks, in particular the differential equations that are used and solved in Black-Scholes theory. When we use the continuous time framework we write

$$\frac{dS}{S} = \mu dt + \sigma dz,$$

from which we can derive the Black-Scholes price for options. This sigma is for a continuous-time process and is different from the finite-time-window sigmas we use elsewhere in this paper. This is the same as the difference between an annual return of, say 5%, which causes $1 to go to $1.05 after a year, and the equivalent continuous compounding return of ~4.879% [= $\log_e(1.05)$] which, with continuous compounding, takes $1 into $1.05 after one year. We note that in practice many people, even professionals, use these finite-time-window sigmas measured through data incorrectly as the continuous-time sigma in Black-Scholes.

The relationships between the finite-window returns, sigmas (R_n, σ_n) to the continuous returns, sigmas (μ,σ) are

$$R_n = e^{n\mu}, \quad \sigma_n = \sqrt{e^{2n\mu}\left(e^{n\sigma^2} - 1\right)}.$$

Lastly, we should also point out that there is another set of notation in common use, that is different from the notation we are using in this paper. The space we are working in is price space *S,* and we are talking about returns where the price-return relationship is

$$R_i = \frac{S_i - S_{i-N}}{S_{i-N}},$$

which is a discrete-time return as discussed in the previous paragraph. Other authors, primarily academics, work in the log space of price ratios, where

$$r_i = \log\left(\frac{S_i}{S_{i-N}}\right),$$

where the return is defined as the log of the ratios of the price. For small time intervals and small price movements these two return calculations are close to equal. In this log space the returns are approximately normally distributed, the arithmetic average is the correct calculation (which is geometric in nonlog space), and the sigma does scale exactly as square root time. In the price space, the notation we use, both the price and the return distributions are lognormal and the sigma scales in the more complicated way involving the return also. The log space notation is much easier for analytic calculations in the random walk, lognornmal assumptions, but it is less transparent than real space. In the real space that we use in this paper, it is more natural to use price data and discrete time steps, and the annualized returns are more natural instead of the continuous time return. This real space is the notation most practitioners use.

We also point out that when we talk of calculating averages of returns, we are generally talking about averages across time. It is in this context that we talk about arithmetic or geometric averages. Another consideration for average returns is a point-in-time average across the distribution - distributions we will show later (e.g., Figure 8). Here we can calculate the mean [integral of R times P(R)], or the median (the 50% probability point in these figures). When the distributions are close to symmetric around the 50% point, the mean and medians are close to the same. We talk about the mean and sigma when describing the distributions, but sometimes think about the median and other locations of expected return, such as a return with 10% probability of achieving or exceeding, when we investigate the risk.

3. Markets Are Not Random Walk

A Random Walk (RW) process is one where the value at each next step has no dependence on the values at previous steps, and the step sizes are all drawn from the same distribution, or Independent Identical Distribution (iid). This is also sometimes called a Markov process, although strictly speaking it is a zeroth-order Markov process. A first-order (second-order) Markov process has steps whose values can depend on the previous (two) value(s). Brownian motion, also called a Wiener process, is a random walk process where the distribution of the step sizes is a normal distribution. A Geometric Random Walk (GRW) is one where the log of the values follows a random walk, that is the relative changes are a random walk. Geometric Brownian Motion (GBM) is a GRW process where these relative changes are drawn from a normal distribution, so the new values come from a lognormal distribution. Frequently people loosely only say random walk when they are describing a RW process with lognormal distribution, or GBM. Data shows that the short-term returns r_i on stock prices have an approximate normal distribution (or log normal for R_i), so stock returns approximately follow a BM process (RW with normal distribution), and stock prices follow a GBM, or a GRW with a lognormal distribution. One property of RW processes is that the stochastic deviation from the starting point moves outward as the square root of the number of steps, or time for stock motion. Therefore this square-root scaling is one test for a RW process.

Everyone is familiar with the classical model of stock motion, loosely called random walk, with a normal distribution of returns for small time steps, or a lognormal distribution of prices. This set of assumptions allows tractable theoretical analysis and beautiful closed-form solutions, such as the Black-Scholes option price model. Everyone also knows that this is just an approximation and that stock motion is not really a random walk with lognormal prices. There exist many extensions to the simple random walk with lognormal prices theoretical framework, still using the analytical approach. Most of the focus of these analytical approaches are on the dynamics of very short-term motion, typically using a continuous-time approximation.

There has been much measurement and documentation of how return distributions differ from normal (see, for example Gopikrishnan[1999] and Farmer[1999]). The focus of these measurements has been on high-frequency returns so that enough data can be collected to investigate the tail behavior, with particular emphasis being on daily or intraday returns. Data shows that short-term returns have distributions with fatter tails than described by normal distributions, with the tails showing approximate power law behavior. Among other things, understanding tail behavior is critical in managing risk and pricing derivative products. It can be shown that any power law tail implies an infinite price for equity options, unless a reasonable cutoff is assumed. There has not been nearly as much emphasis on long-term return distributions since the data is much more sparse.

An interesting exposition on the stochastic nature of markets was recently put forward by Mandelbrot (2004), who has been a long-term proponent/inventor of fractals. He strongly believes that markets exhibit a fractal nature, where the dynamics are scale-independent, meaning that the behavior looks the same regardless of the scale one examines. While this certainly looks true qualitatively, it is not clear what the practical implications of such a diagnosis is, except to further discredit the random walk market behavior.

A recent book by Andrew Lo (2002) summarizes a number of statistical tests through the years to address the RW nature of markets. His conclusions are that markets do indeed exhibit substantial deviation from RW, but it is inconclusive about how much advantage can be used for trading. As we will show in the following sections, long-term stock return distributions exhibit strong mean reversion, with substantial consequences for long-term strategies such as asset allocation.

4. Empirical, Data-Driven Approach to Market Analysis

As discussed in Section 1, the standard approach to modeling market behavior is a combination of theory with statistical properties calculated directly from the historical data. This is generally considered a robust approach since the use of theory to project future behavior creates independence from specific market realizations in the past, or so it is argued. However, there are two fundamental problems with this approach:

(1) The theory does not match the data for either long or short term behavior, and

(2) The most sensitive statistical parameters, the mean return and sigma, are measured from a finite set of data that has the "only one realization" problem.

The approach described in this paper is based on a completely nonparametric, data-driven viewpoint where we make no *a priori* assumptions about the distributions or the evolutionary dynamics. We believe that this empirical, nonparametric data-driven approach is a more realistic, more accurate approach than using the standard combination of imperfect theory with "only one realization" data. Both approaches, indeed any approach, suffer from the argument that we have only seen one representation of all possible paths in the past, and require the assumption that the future will, at least stochastically, behave similar to the past.

4.1 Long-Term Equity Market Performance

We examine the historical evolution of the U.S. equity market over as long a time period as is possible, and that we believe has relevance to today's markets. Figure 6 shows the total return of large company stocks from January 1926 through December 2003, and covers a time frame with many important and relevant events, including two bubbles, a depression, several recessions, strong bull and bear markets, and several crashes of varying suddenness, depths and durations. Again, we argue that the behaviors directly exhibited in this data are more likely to statistically represent the future than a combination of two statistical quantities combined with theory.

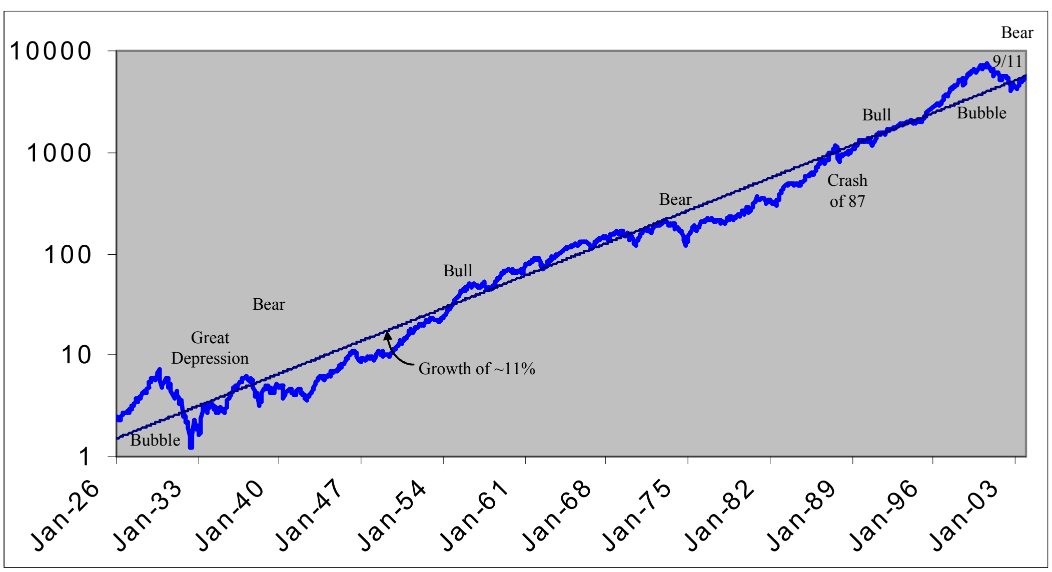


Figure 6. Historical Total Returns for Large Company Stocks Since 1926.

The straight line in this picture shows a steady exponential, continuously compounding growth of about 10.5%, and visually shows the approximate average growth rate over this 78-year period. The data is on a monthly frequency, and represents an index of large company stocks, similar to an S&P index, with all dividends being reinvested as they are earned.

In this picture we see that the largest event by far was the bubble leading up to, the crash and subsequent brutal period of the Great Depression. We can see several strong bull markets and several long bear markets, periods of many years where the average returns over, say, a decade were significantly over or under 11%. It's interesting to put such large effects as the crash of 87 into long term context, where it's hard to identify as the significant event it was. Closer to home we can clearly see the irrational exuberance of the second half of the 90's followed by the bursting of the dotcom bubble.

This picture of U.S. stocks is similar to what has been experienced in equity markets in modestly mature countries outside the U.S. In Jeremy Siegels's *Stocks for the Long Run* we find long-term equity markets of the U.K., Japan and Germany showing similar long-term behavior as the U.S., albeit each with their own idiosyncrasies. Further, in this same reference we can find data going back to 1801 with strong evidence of U.S. equity market solid performance with no obvious substantial long-term statistical differences in these two centuries.

4.2 Probability Distributions of Long-Term Returns

We use the data shown in Figure 6 to calculate a variety of statistical properties around long-term market performance. We follow the process described in Sections 2.1 and 2.2 to directly calculate the multiyear returns by looking at all possible windows of the desired duration over this 78 year period.

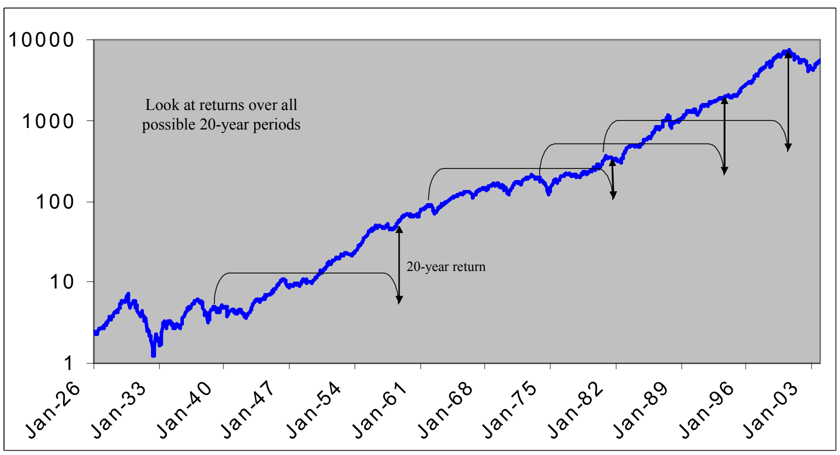
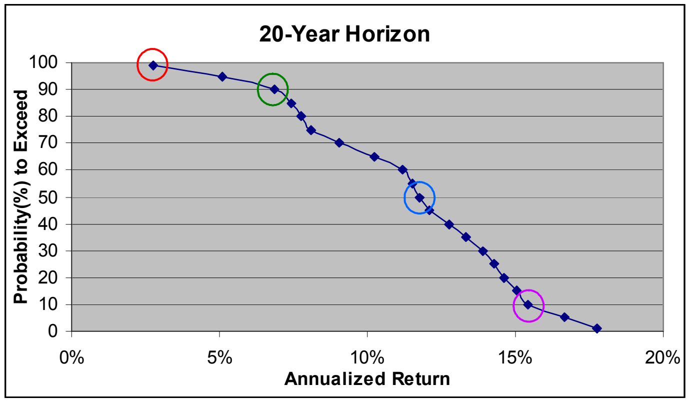


Figure 7. Use All Possible 20-year Overlapping Windows to Create the Statistical 20-Year Return Distribution.

Figure 7 shows an example of how we directly calculate the 20-year real returns using all possible 20-year time windows, including all overlapping windows. As we argued in Section 2.1, this overlapping window process only improves over the ~4 nonoverlapping 20-year windows during this 78 year period. Further, we submit that, even with only 4 nonoverlapping 20-year periods in our data, this direct, empirical, nonparametric estimation of the 20-year return distribution (and corresponding sigma) is a more accurate and relevant estimate of the dynamics of real markets than would be a shorter term (e.g., one year) average of return and sigma, projected to 20 years using random walk theory.

From this collection of all possible 20-year returns we directly calculate the mean, median, standard deviation and the cumulative distribution. We use the cumulative distribution since it is more robust to statistical inference than the point distribution. We scale the return to an equivalent annual return that produces the 20-year return.



Holding equities for 20 years,

- 99% chance to exceed 2.7% annualized return
- 90% chance to exceed 6.9% annualized return
- 50% chance to exceed 11.8% annualized return
- 10% chance to exceed 15.5% annualized return

Figure 8. Cumulative Probability Distribution of 20-Year Equity Returns.

The indicated data points in this figure are at each 5% interval in the cumulative distribution, except for the two endpoints at 1% and 99%. From this figure we can directly estimate the probability of achieving a particular performance over a 20-year holding period. We find that over a 20-year holding period, we can exceed a 2.7% annual return with 99% confidence, and we have a 10% chance of exceeding a 15.5% annual return. The median return is 11.8% over all 20-year equity holding periods since 1926.

As discussed above, we should be careful about this 20-year holding period distribution. It is formed from about 80 years of data, so it has about 4 nonoverlapping periods, and we then use the overlapping windows to smooth it. We have reasonable confidence in the mean and the overall basic shape of the distribution, particularly the spread, but the specific “kinks” should not be considered real; they are likely not intrinsic to these 20-year returns and will likely not be observed in the future. Therefore, while one should consider mentally smoothing Figure 8, we believe that the overall features (mean, spread, and to some extent the tails) are statistically robust and are the most accurate way to estimate stochastically into the future. There is no reason to believe that a projection of one year data out to these long holding periods using a theoretical model would provide a more realistic distribution than the one measured in Figure 8.

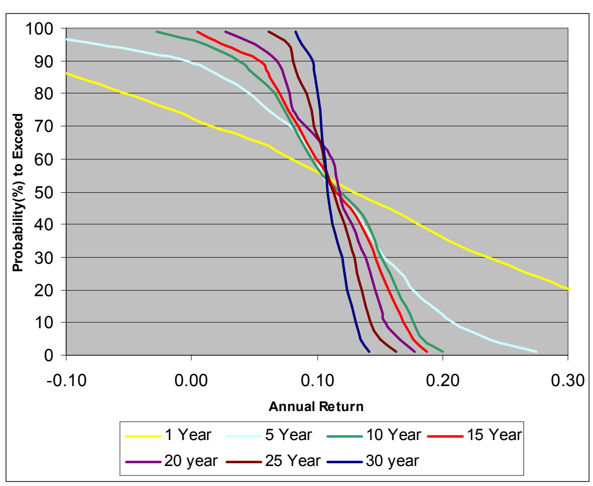


Figure 9. Cumulative probability distributions for different time horizons for holding stocks. Each curve gives the probability to exceed a particular return for that holding period, as calculated from data from 1926 through 2003.

In Figure 9 we show a family of these cumulative probability distributions for a variety of holding periods up to 30 years. These returns are calculated using the sliding overlapping windows method and use a window size of the particular holding period. The returns are then annualized by computing the annual return which, with compounding, will give the n-year return. This is done through the formula

$$R_1 = (1 + R_n)^{1/n} - 1,$$

where R_n is the n-year return from the window size of n years, and R_1 is the associated 1-year return, which if compounded annually for n years, will give the return R_n.

Another way to plot this data is shown in Figure 10 below. Here we plot the expected annualized return at different confidence levels versus holding period. A similar figure is shown in Ibbotson, but in that publication they use the 1-year values of return and sigma, and then project them to multiple year holding periods using a random walk assumption, where the mean return is constant and the standard deviation scales approximately as

$$\sigma_n = \frac{\sigma_1}{\sqrt{n}},$$

which assumes the random walk iid process. The result of this theoretical projection is shown in Figure 11.

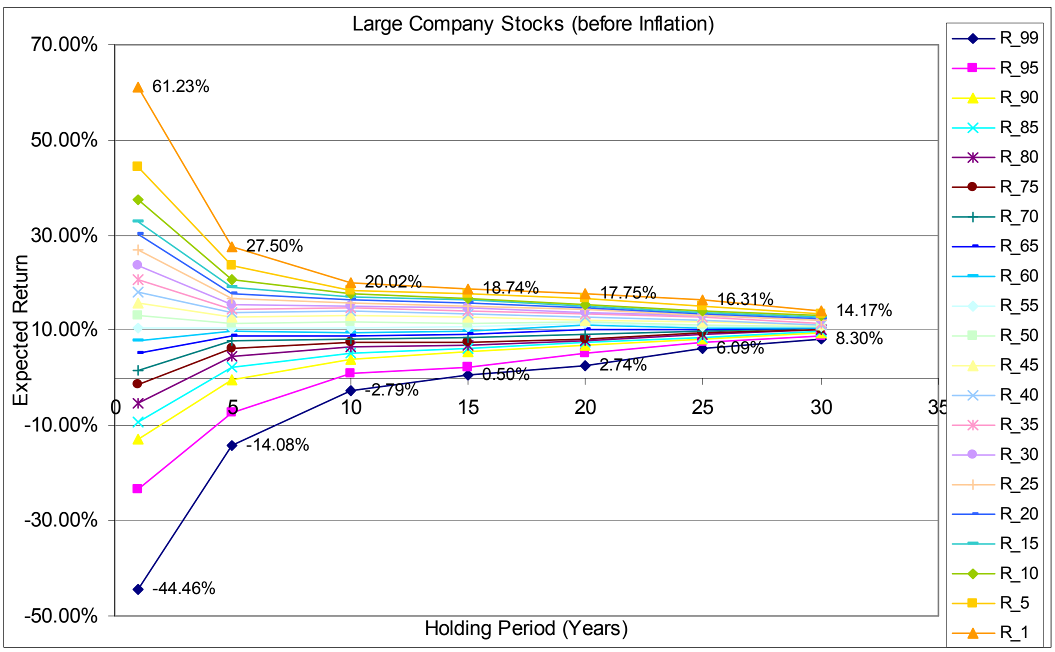


Figure 10. Expected annualized return for stocks at various confidence levels for different holding periods. This figure is generated using actual historical market data.

We can compare this Figure 10 with the result generated using the random walk assumptions. In Figure 11 we show the identical information as Figure 10, but in Figure 11 we first fit a normal distribution at holding period one year using the mean and sigma at one year, and then project out this distribution in time using the random walk assumption (mean return is constant, $\sigma_n = \sim\sigma_1/n^{1/2}$).

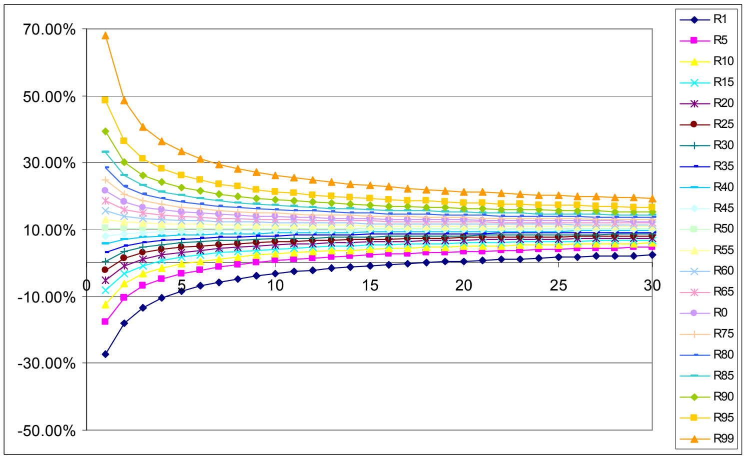


Figure 11. Expected return for stocks at various confidence levels for different holding periods, using a random walk model. This figure is generated using actual one year mean return and sigma from data, and then using a random walk model to project forward in time. This random walk projection is noticeably different from how the market actually behaves.

We can immediately see in comparing Figure 11 to Figure 10 that the standard deviation of returns, as measured directly from the data, is less than that from a random walk assumption, particularly as the holding period increases. From the actual data distributions, Figures 9 and 10, we can compute a sigma without assuming either a random walk or a normal distribution, through the standard sigma formula

$$\sigma = \sqrt{\frac{\sum (R_i - R_{avg})^2}{L-1}} .$$

With this formula we use the actual multiyear return data to calculate a mean annualized return and sigma of the annualized return directly from the data, and we can compare this to the square root time scaling of sigma from the random walk model. This comparison is shown in Figure 12.

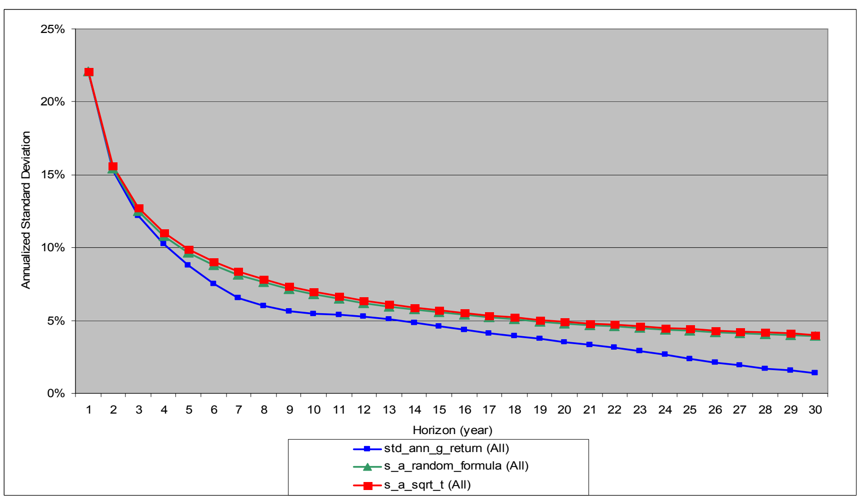


Figure 12. Standard deviation (sigma) of equity annualized returns as a function of holding period. The top two curves are from using the one year value and then projecting using a (1) a square root time scaling or (2) the more correct scaling given in Table 1 using both the annual sigma and return. There is no significant difference between the approximation (1) and the more accurate formula (2). The bottom curve shows the sigma calculated from actual data, not using a random walk, normal distribution assumption, and shows substantial deviation from the theory.

This comparison of the theoretical scaling in sigma to the actual data is performed frequently and is often referred to as the variance ratio test [see, for example, Andrew Lo]. This test, equivalent to investigating the autocorrelation function, is usually used to investigate short term deviation from random walk, looking at relationships on the order of minutes up to about a year. Here we are looking at very long term behavior, from one to thirty years, and we find evidence of substantial deviation from random walk.

4.3 Equity Markets Show Strong Mean Reversion and Skinny Tails

Following the observation of Poterba [1988], we agree the evidence shows that equity markets have positive autocorrelation for short times and negative autocorrelation for long time horizons, which says that trends exist for short time periods and mean reversion exists for long time horizons. They also find that there is even stronger evidence for mean reversion in the data from 1871 to 1925 than from 1925 to 1987.

It is well known that short-term return distributions show fat tails - tails that fall off substantially slower than a normal distribution, as discussed in Section 3. In this report we are studying the shape of long-term return distributions, for holding periods from 1 to 30 years, using these explicit holding periods directly and measuring these distributions nonparametrically. We measure the returns using these long-term jumps, annualize these returns, and then investigate the distributions of these n-year holding periods, comparing them to what a random walk model would give projected from the same 1-year holding period starting point.

Figure 13 shows the cumulative distributions of these long-term equity returns (green curve) compared to the random walk projection (blue curve). This is the same data as shown in Figures 10 and 11, but now put together for different holding periods.

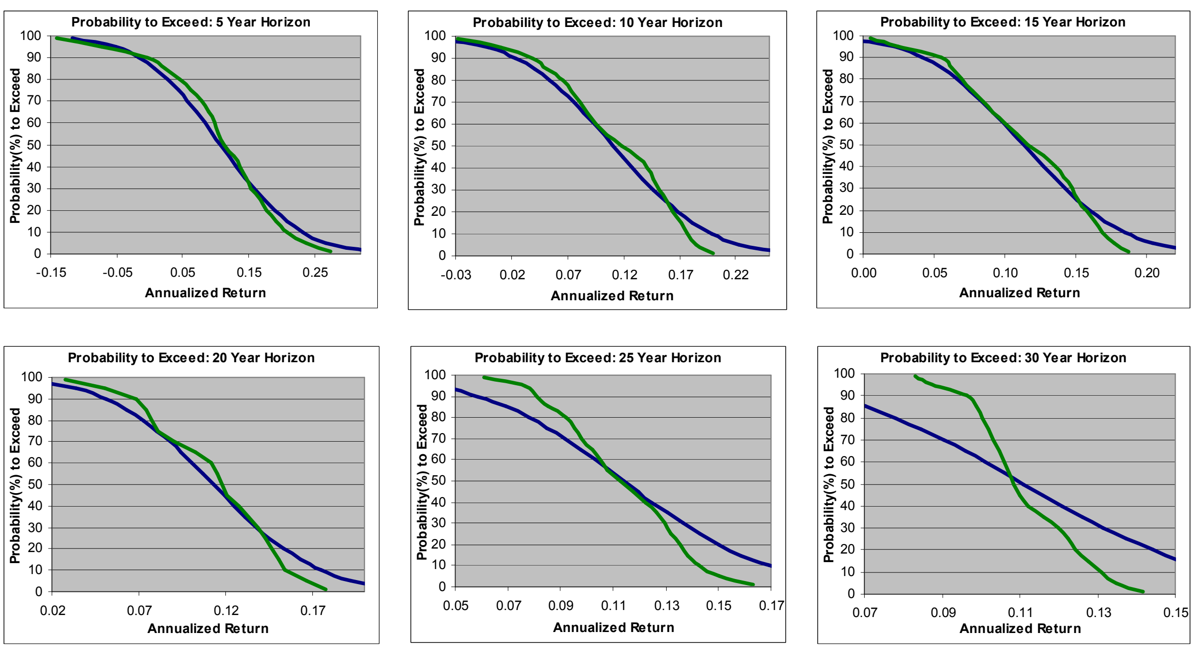


Figure 13. Comparison of the cumulative distributions of multi-year returns using (1) one year measured values, normal distribution assumption and random walk projection (standard approach, blue curve), and (2) nonparametric measurement directly from the data. Substantial differences are observed, particularly for long-term holdings.

These pictures show that there is a striking difference in the long-term return distributions between theory and data. This behavior is not unknown, and is also described in Siegel, Bogel, and others. In this paper we quantify it and describe a good methodology to empirically calculate the distributions.

A few observations about what the data shows as compared to theory:

- We know the short-term (on the order of a month or less) distribution has fat tails, both high and low (positive and negative returns). The 5-year distribution still shows fat tails on the low side but already has skinny tails on the high side.
- At 10 years and higher we see both the high and low tails are skinny.
- The effect of the smaller sigma is seen as a much more narrow distribution, particularly from 20 years and beyond.
- The risks of stocks are substantially less than theory suggests for holding periods of 20 years and greater.

The impact of these data-driven conclusions is significant for any long-term analysis, particularly for asset allocation.

In examining the behavior of the tails of the long-term stock return distributions we can also ask how different they are from a normal distribution using a sigma appropriate for that distribution. We have seen that sigma scales much differently than square root time, but how much different is the distribution from normal if we use the actual sigma? We perform this analysis, shifting the normal distribution to be centered on the empirical distribution using a least squares fit, and then closely examined the tails in the empirical distributions compared to normal distributions using the correct sigma. What we find is

- Holding period 5 years: skinny tail on high side, fat tail on low side
- Holding period 10 years: skinny tail on high side, fat tail on low side
- Holding period 15 years: skinny tail on high side
- Holding period 20 years: skinny tail on high side
- Holding period 25 years: skinny tails on both sides
- Holding period 30 years: skinny tails on both sides.

These statements should be taken with some caution since the tails of the distributions are sensitive and difficult to analyze. However, we believe they do represent a realistic deviation from normal distributions, and suggest that even with the more realistic sigmas, the long-term distributions exhibit skinny tails.

As we described before, as one goes out to these long time windows the accuracy of the distributions lessens. What then is the best way to estimate a 30-year holding period return distribution? At some point one may lose confidence in the accuracy of these empirical distributions, even though the use of overlapping windows provides smoothing. We mentioned before a heuristic of requiring at least a few nonoverlapping periods, so our confidence in the 30-year window distribution may not be that high. However, we should have high confidence in the 10 or 15-year holding period distributions, since they have 5 to 7 nonoverlapping periods. If one loses confidence for periods longer than this, a good method of estimating these higher holding periods would be to begin with a holding period that one has confidence in, say 10 years, and then use theory to project from there. Once could take the 10-year distribution shape shown in Figure 13, and then use the theoretical square root time scaling to project it to 30 years or higher, keeping in mind that the data shows the sigma will likely shrink faster than the theoretical scaling. We believe that the direct, empirical method used in this paper will give a more accurate representation for the 30-year distribution than (1) a theoretical projection from one year data as well as (2) a theoretical projection from the 10-year distribution in Figure 13. However, for those skeptical of the accuracy of the 30-year distribution given here, it's difficult to argue about the compromise of using a theoretical projection from the 10-year distribution. This 10-year projection must be better than a one-year projection, and it will have substantial impact on asset allocation. If we use a square root time scaling starting from the 10-year distribution and put it in Figure 12, we find that there is substantial further mean reversion occurring from the holding periods between about 18 to 30 years, which has a dramatic effect on asset allocation. We chose to rely on this effect as measured in the data as opposed to relying on the theory, because we already see significant departure from theory from years 1 to 10. Inspection of the longer term returns from 1801 in Siegel's book also increases our confidence in the stability of this long-term process.

It is not hard to understand and explain both short-term and long-term market behavior in terms of investor psychology and free markets. Short-term trends exist, and can even become disastrous bubbles. Long-term mean reversion comes from business cycles, driven by the need for substantial over- or under-pricing to be washed through the system over long time scales.

5. What is Risk?

There has been and will continue to be much discussion on what is risk in the financial world (see, e.g., Bernstein). The classic definition of risk is the standard deviation or sigma of the distribution of outcomes, which is equivalent to uncertainty. A related but different family of risk definitions is built around the likelihood of undesired results, such as Value at Risk (VaR) and Shortfall Risk (Liebowitz). All of these approaches have merit and use, depending on what one is trying to measure. The use of uncertainty (sigma) as risk is a natural choice and is easy to understand and employ. However, the use of sigma as a measure of risk has large issues when (1) distributions are not normal or symmetrical, (2) when comparing distributions with very different means, and (3) when looking at long-term compounding. Further, the use of sigma as a risk measure fails to differentiate between downside, undesired results and the possibility of achieving unusual highside results. VaR and Shortfall Risk have nice intuitive meanings, built around the expected losses with low but nonzero probabilities, and are designed to focus on the low-side tail events. In particular, Shortfall Risk explicitly examines the probabilities of achieving results above/below chosen benchmarks, such as what's the probability of achieving a return above, say, -5%. This is the framework we use in constructing the cumulative distributions above (Figures 9 through 13).

The figure below shows four different pairs of portfolios, showing the cumulative distribution of expected return similar to the previous pictures. These four examples are designed to provoke a discussion on what is risk and how to evaluate the risk of various portfolio choices.

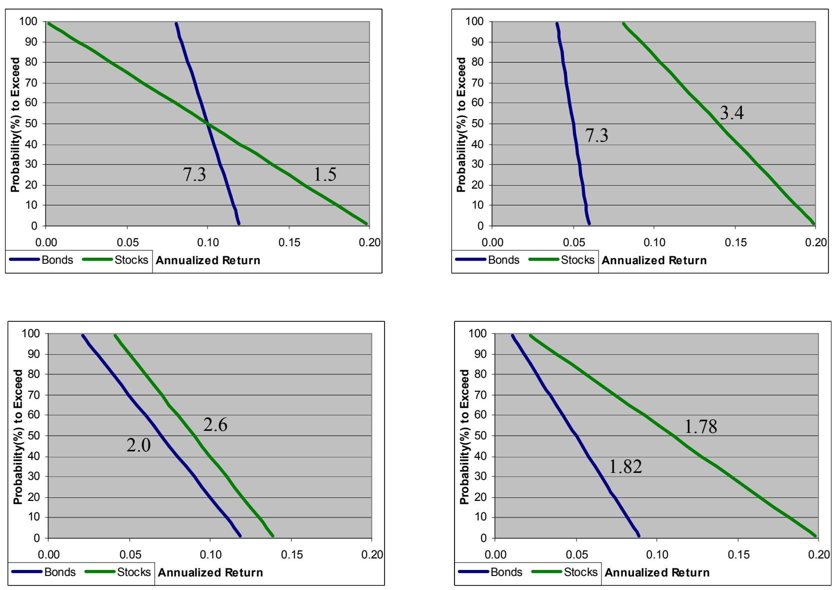


Figure 14. Schematic, side-by-side comparisons of pairs of different portfolios. The numbers shown are the Information Ratios (Sharpe Ratios with risk free rate set at zero). In each example, which is the riskier of the two portfolios?

In the first figure above we see the two portfolios both have the same expected value of return (50% probability) of about 10%. However, the green portfolio has a much larger sigma (uncertainty) than the blue, and by the traditional measure is therefore much riskier. We concur with this assessment, and even the Information Ratios (IRs) do well at identifying the riskier portfolio. We would say that, by any reasonable measure, the green portfolio is riskier than the blue in this first picture.

The situation is not as clear in the second, upper right picture. Here, while the green portfolio has a higher uncertainty (sigma), it in all cases performs better than the blue portfolio. We should conclude that the blue portfolio is riskier than the green, since the green always outperforms the blue. Neither the sigma nor the IR lead to the identification of the green curve as being less risky, and these standard risk measures would have us choose the blue portfolio over the green as less risky, even though it is guaranteed to lose to the green portfolio.

In the third (bottom left) picture we see the distributions both have the same sigma, but the green is shifted to the right from the blue. Here we should agree that the green is the better portfolio, and less risky in the sense that it generally outperforms the blue portfolio. The sigma doesn't point to this conclusion but in this case the IR does, being higher for the green over the blue. We note in this picture that the distributions overlap over a range of returns. This means that, as opposed to the second figure, in this third case it is possible for the blue to outperform the green portfolio. We see, for example, that there is about a 20% chance for the blue portfolio to exceed a 10% return, and there is about a 60% chance for the green to underperform 10%. Because the curves have overlap it is possible, but less likely, for the blue to outperform the green.

What do we choose in this case? We can reason that for any target return the green curve always has a higher probability of hitting or exceed it. Therefore the green is always a better choice in this example, and will always have the expectation of having a higher return. By this reasoning the green curve is less risky than the blue curve for this third case. We recognize that our decision criterion has become

- Whenever one curve is always above and to the right of the other, the curve on the right is the better portfolio and less risky, since it is always expected to perform better than the one to the left.

This doesn't mean that it won't underperform the one to the left, but that the expectation of the stochastic process with the distribution to the right and above is always better than the one to the left and below. In this analysis we assume that the two portfolios are uncorrelated. The conclusion is still the same regardless of the degree of correlation, but we note that in the case when they are 100% correlated, this particular green/blue pair would then be a perfect arbitrage situation.

By this logic, when we examine the fourth picture, we should identify that the blue curve is more risky than the green curve since it is always expected to underperform the green curve. We note that again the standard definition of risk, both sigma and IR, lead to the opposite conclusion for this pair. We believe that the rule in the bullet point above is, in general, a better decision rule in identifying the more risky portfolio. It is important in using such a rule to take into account one's belief in the constructed distributions, and to closely examine the lower-side tail of the distributions.

In most cases the decision is not this simple. Usually the two portfolios overlap as we will see below. Second, the portfolios are obtained from a finite number of data points (an issue for all financial modeling in any approach), and we should be cautious about trusting the distributions completely, especially the tails. If we had 100% certainty in the distribution the above stated rule is completely and always correct, but still doesn't address cases of crossing. Given uncertainty and the fact that frequently the left-hand tails of the distribution cross, we should exercise judgment with regard to our confidence in the left-hand side tail behavior of both distributions in making risk assessments.

6. Examine the Long-Term Returns of Bonds

We now examine the performance of bonds using the same methodology as we did for stocks. We use a time series of total return from long-term triple- and double-A rated corporate bonds. The data is shown in Figure 15, which also shows the stock data used above as well as CPI. The bond data is representative of a long-term, high-quality corporate bond fund.

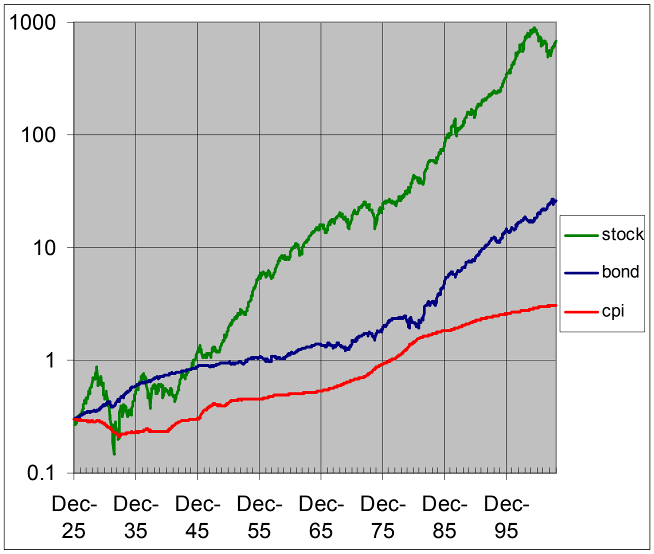


Figure 15. Total returns for stocks, bonds and CPI, with arbitrary normalization so they are equal at 12/1925.

We do all the same calculations as we did for stocks, using discrete long-time-horizon overlapping windows, to build all the cumulative distribution functions of returns. Figure 16 shows these distributions for many different holding periods.

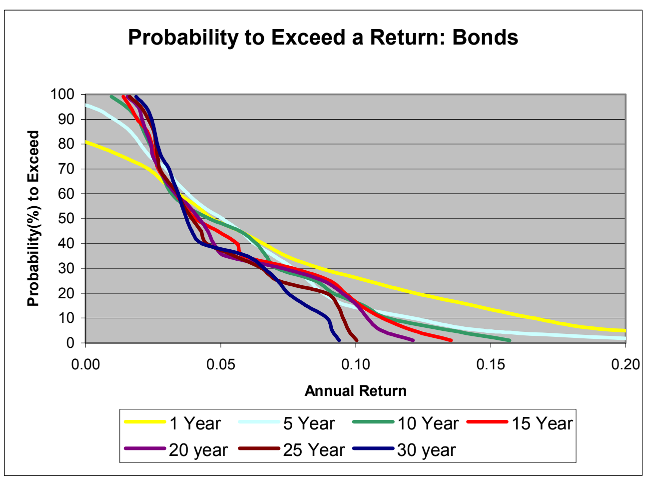


Figure 16. Cumulative distributions functions of annualized returns for bonds for different holding periods.

These distributions can be plotted in a different way, as in Figure 10 for stocks.

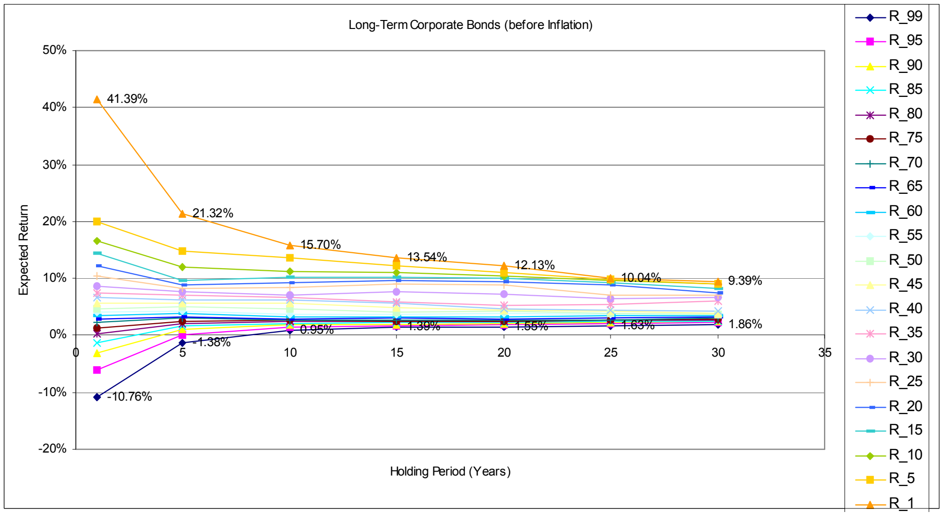


Figure 17. Expected returns for bonds at various confidence levels for different holding periods.

As before, we can examine the time dependence of the sigma of this annualized distribution. If it were a random walk, sigma should decay as square root time, but here we see a very interesting phenomenon. Whereas for stocks (Figure 12) we saw sigma shrink faster than random walk theory, for bonds we see the opposite. This implies that, for the period of time we are investigating (1926 - 2003), bonds exhibit the opposite of mean reversion. This is perhaps an exhibition of substantial trends, particular at long time scales, up to about 20 years. After about a 20-year holding period the sigma starts to decay faster than random walk. We might then conclude that there is substantial momentum in bond returns up to about a 15-year holding period, after which there is mean reversion, at least for the time period from 1926 through 2003.

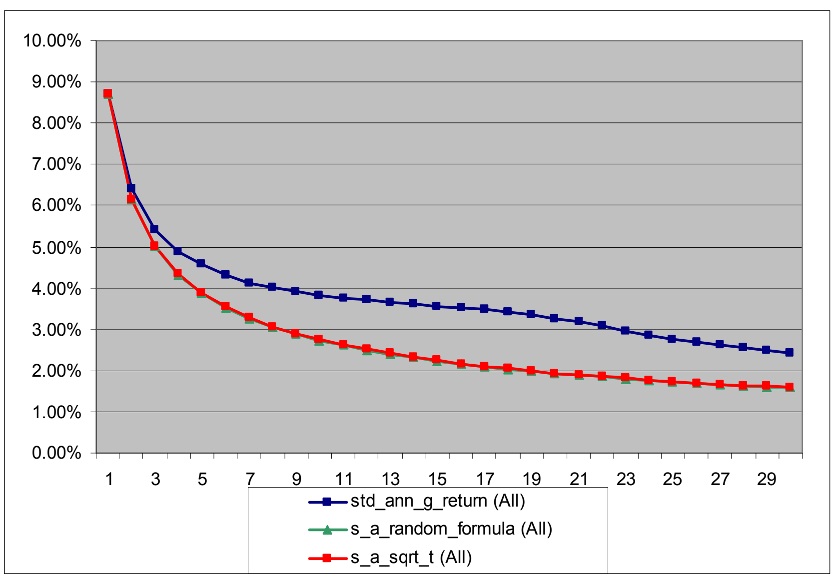


Figure 18. Standard deviation of bond annualized returns from actual data (blue) compared to a scaled using the usual theory. The behavior is opposite of stocks in that we see mean AVERSION (trends) until a holding period of about 15 years, after which we see mean reversion. Data from 1926 through 2003.

In this figure we notice that the simple square root time scaling is even closer to the more correct formula described in Section 2.2 (the red and green curves), since for bonds both sigma and the return are closer to zero than for stocks.

6.1 Bonds are Riskier than Stocks for Long Term Holdings

We can compare the sigma of bonds to stocks for different holding periods, shown in Figure 19. For both instruments we also show the theoretical scaling, shown in red.

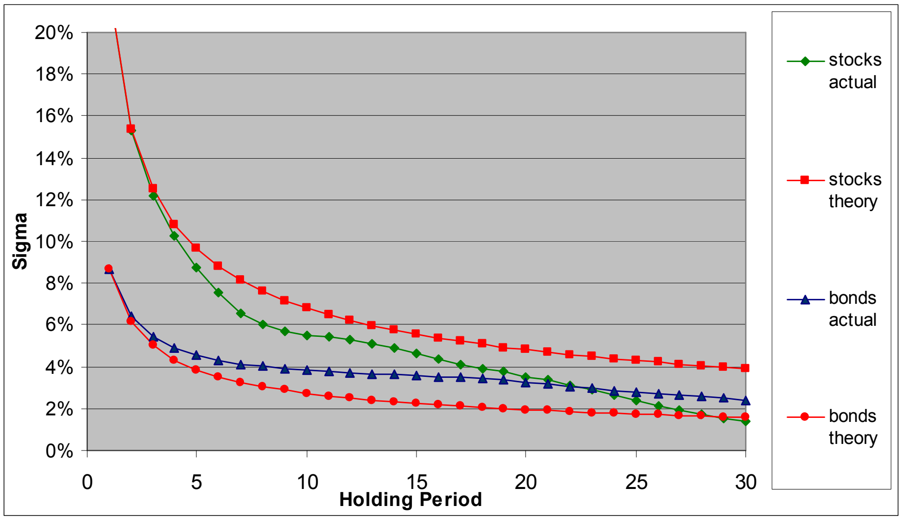


Figure 19. Reduction of sigma as the holding period increases for both stocks (green) and bonds(blue), as well as the scaling using random walk with lognormal distribution theory (red).

From this figure we can see the unexpected effect that, not only does sigma for stocks/bonds decay faster/slower than theory would say, but that the sigmas for stocks and bonds actually cross at a holding period of around 20 to 25 years. This says that, by the traditional measure, bonds are more risky than stocks for holding periods greater than about 20 years, which is an unusual statement.

We continue to examine the relative performance and risk of stocks and bonds by looking at the short and long-term behaviors. Figure 19 shows these comparisons for one- and 30-year holdings.

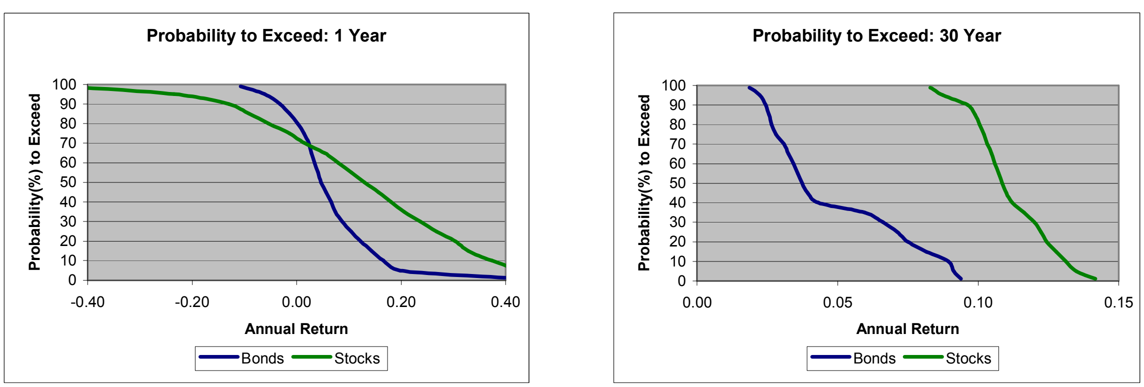


Figure 19. Comparison of stock and bond performances for short (one-year) and long (30-year) holdings.

The first picture shows nothing unexpected. We see that, although stocks have a higher expected return for a one-year holding, the uncertainty (sigma) is substantially higher for stocks, and further the left side of the distribution shows substantial likelihood of stocks underperforming bonds. For these reasons it is quite clear that stocks are much riskier than bonds by any measure, for a one-year holding.

The second picture shows the opposite. By any measure, bonds are more risky than stocks for a 30-year holding. Not only is the traditional risk measure sigma higher for bonds than for stocks, but the stock distribution is substantially to the right everywhere of the bond distribution. Finally, in examining the left-hand-side tails in the 30-year holding distributions, there is no evidence that the tails in stocks will be lower than those for bonds in the future.

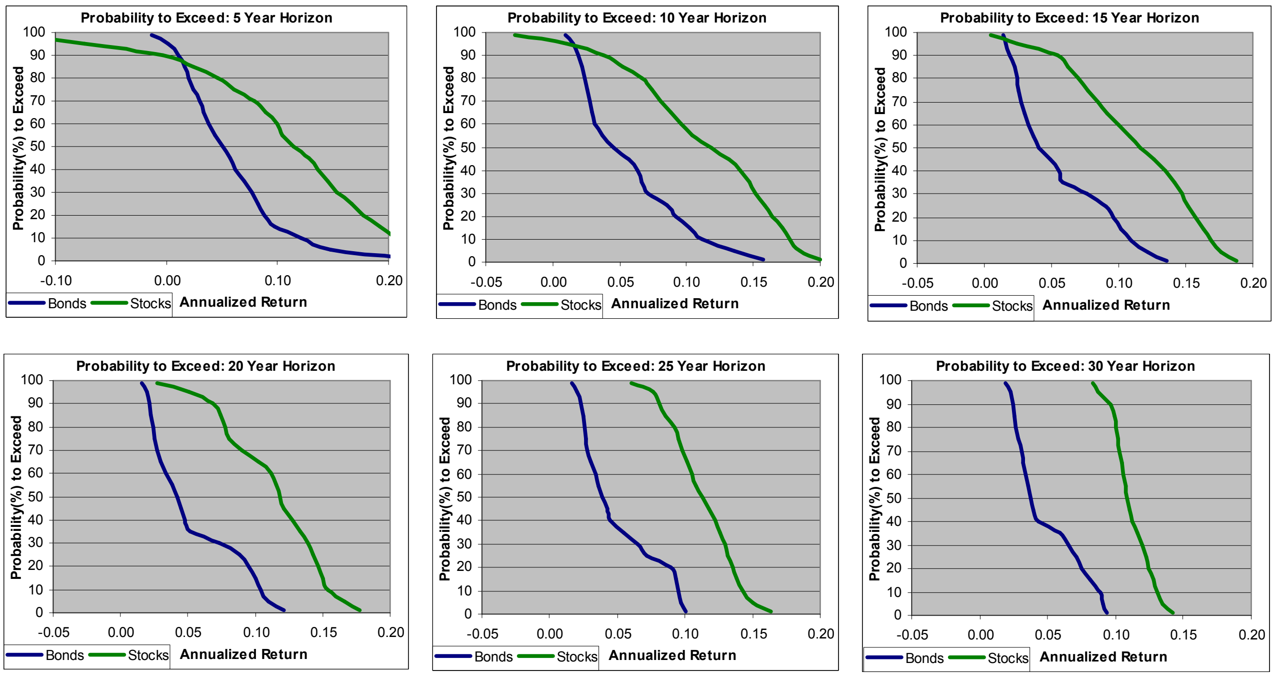


Figure 20. Comparison of the distribution of annualized returns of stocks to bonds, over different holding periods. Bonds are riskier than stocks for holding periods about 15 years or greater.

In Figure 20 we show the comparison between stocks and bonds performance for six different time horizons. In the first, 5-year holding picture, we see that stocks almost always beat bonds, but in examining the left hand tail, which is the important consideration from a Shortfall Risk point of view, we see that there is still a sizeable chance (about 10%) that stocks will perform worse than bonds, and when they do they can be much worse. Even though the stock curve is usually to the right of the bond curve, we would still conclude that, by the Shortfall Risk perspective, stocks are probably more risky than bonds for a 5- to 10-year holding period, particularly if 5% events are important to you.

However, beginning around the 15-year picture and continuing, all the longer holding periods show that stocks are always expected to perform better than bonds for holding periods of greater than 15 years, to the 1% confidence level. Not only are stocks better, but they are much better and even show lower sigma at higher holding periods. A reasonable reading of this data requires us to be cautious of the distributions, the tilt of the distributions (sigma) and in particular the slopes and separation of the left-side tails in each picture. A reasonable conclusion is that for holding periods greater than 15 years, stocks are less risky than bonds.

7. Data-Driven Approach to Asset Allocation

So far we have described an empirical approach to quantifying the likelihood of achieving particular returns for both large company stocks and bonds, and a heuristic methodology to describe and compare risks of holdings. We now use this approach to investigate asset allocation strategies.

Figure 21 is an enlargement of Figure 10, with emphasis on the area for assessing Shortfall Risk for stocks.

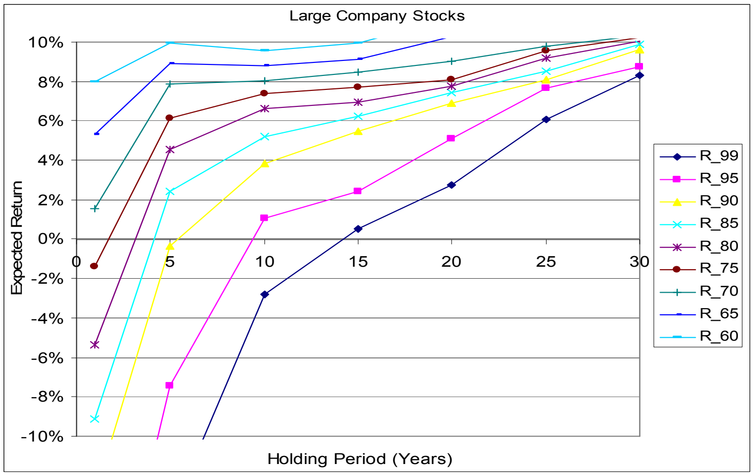


Figure 21. Enlargement of Figure 10 to investigate Shortfall Risk of stocks.

In this figure we can follow any particular probability curve. For example, if we choose the 90% (yellow) curve, we see that we have a 90% probability of exceeding about

- 0% return for a 5-year holding
- 4% return for a 10-year holding
- 7% return for a 20-year holding
- 9.5% return for a 30-year holding.

These curves are very similar to yield curves for bonds.

In Figure 22 we show the same picture for bonds, being an enlargement of Figure 17.

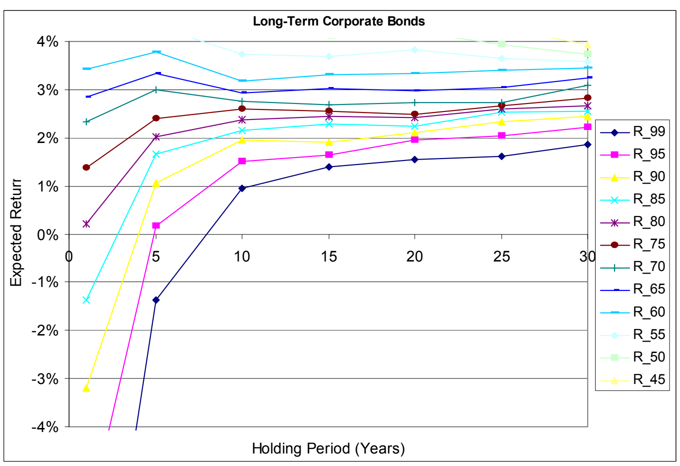


Figure 22. Enlargement of Figure 17 to investigate Shortfall Risk of bonds.

Again, we can think of Figure 22 in terms of probabalistic yield curves if we wish.

As discussed earlier, there are many approaches to asset allocation, and here we consider a different approach. First we ask a very simple question: what would be the optimal initial asset allocation if we were to hold a mixture of stocks and bonds for different holding periods without rebalancing? To answer this specific question we run 9 new analysis of mixed portfolios of 10, 20, 30, …% stocks and 90, 80, 70, …% bonds in addition to the two 100% stocks and bonds we already have. For each of these mixed portfolios we can produce figures similar to Figures 21 and 22, showing the likelihood to meet or exceed a return for the different holding periods.

Using the analysis above we can quantify this answer explicitly as a function of risk tolerance, in the form of selecting a likelihood to achieve. For example, we could ask the above question in terms of likelihood to not miss the goal. Consider what might be called a moderate risk tolerance, that we require a 90% probability to meet or exceed our goal, which is equivalent to allowing a 10% chance that we might fall below our goal. We might then ask, what is the optimal initial mix of stocks and bonds as a function of holding period with a 90% degree of confidence? Figure 23 below shows the answer to this question.

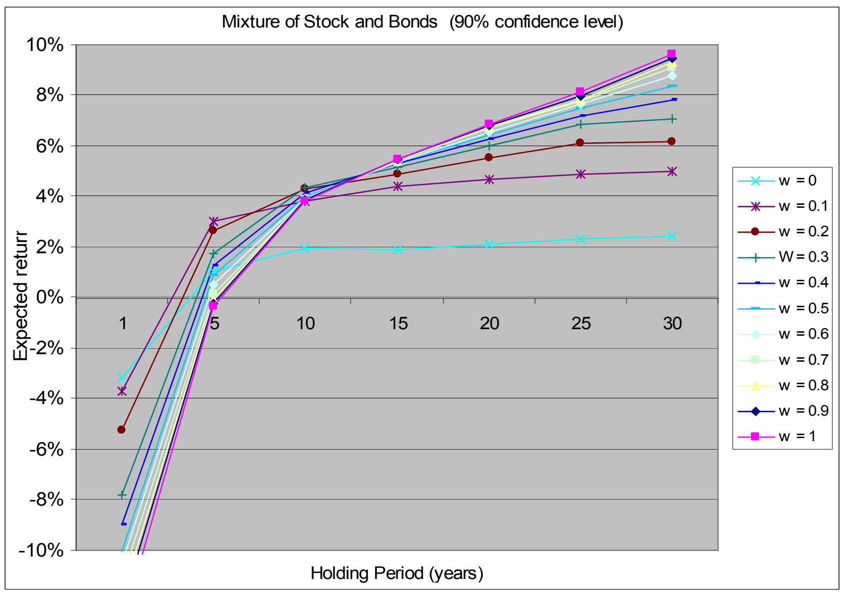


Figure 23. Expected returns with 90% confidence to meet/exceed for different mixtures of stocks and bonds for different holding periods. The parameter w is the fraction of stocks in the mixed portfolio.

Using Figure 23 we can see what we need to do to meet our goal. We should always choose the highest curve for our selected holding period. Therefore the optimal asset allocation with the 90% confidence choice is the upper envelop of these curves. We see that for long holding periods, anything greater than about 15 years, 100% stocks is the best mixture. For holding periods less than about 2 years, 100% bonds is the best. From about 2 to 7 years, 10% stocks is best. The optimal mixture then smoothly increases during the holding periods from 7 to 15 years, going from 10% to 100% stocks. This picture can be made more smooth by using smaller increments than 5 years in holding periods, and in stock/bond fractions. This process will then generate an optimal initial asset allocation percentage that answers the specific question posed above. This optimal asset allocation curve is shown in Figure 24.

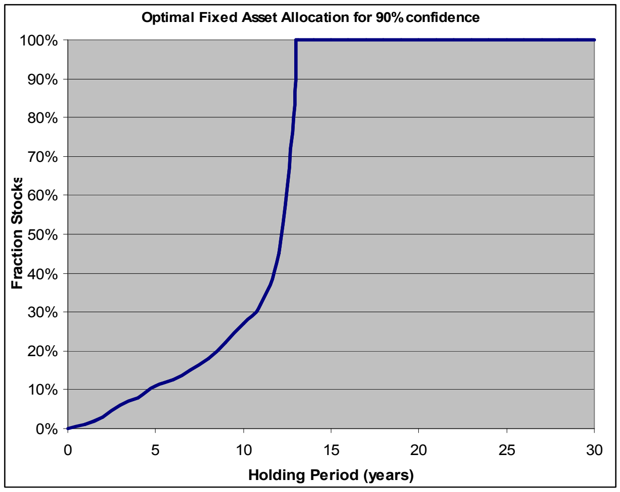


Figure 24. Initial asset allocation to achieve the highest return with 90% confidence for different holding periods with no rebalancing.

It's interesting to note from Figures 23 and 24 that the optimal portfolios for the 90% confidence level are pure bonds before 1 year holding, pure stock after about 15 years holding, and the mixed portfolios do better in between 1 and 15 years holdings. Although we didn't calculate these curves, it's interesting to speculate on what the optimal asset allocation would be at different confidence levels, and this is shown in Figure 25.

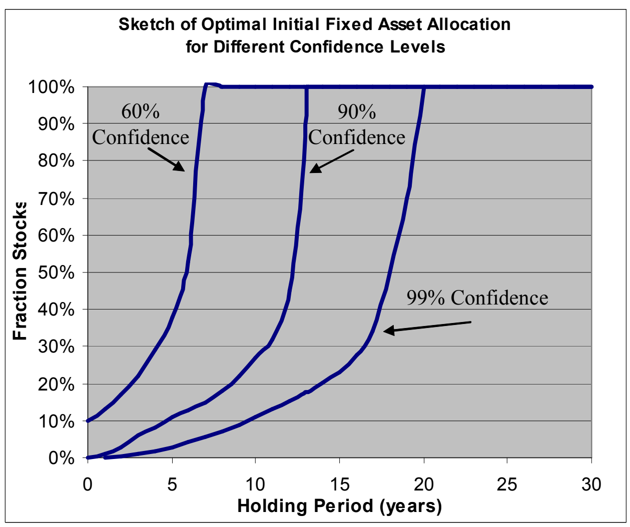


Figure 25. Initial asset allocation to achieve the highest return with no rebalancing at varying confidence levels. The 60% and 99% curves are not calculated; this is an illustration of their qualitative behavior.

In this empirical approach to asset allocation we achieve two new things: (1) we can now explicitly quantify risk tolerance to asset allocation through the Shortfall Risk concept, and (2) we can use the data to simulate any combination of complex, time-dependent rebalancing possibilities and consumption needs. The specific question asked for which Figures 24 and 25 are the answers is a simple and not very realistic question. It is used here as an example of how to answer a particular asset allocation question using a simulation over all known and relevant data. A more realistic question that can be answered with this same process might be

> What is the best time-dependent asset allocation strategy, with rebalancing, to optimize my returns with x(t)% likelihood of meeting or exceeding my goal?

This is a very practical question that not only considers continuous time rebalancing but also a time-horizon-dependent risk tolerance, explicitly given as a likelihood of not meeting one's goals. This question can be answered stochastically using the above-described procedure. Thus we find that with this methodology we can answer simple to complex asset allocation optimization questions with an explicit risk tolerance factor expressed as a probability of achieving the outcome.

8. Philosophy of Data-Driven verses Theoretical Models

Much has been written in the past ~20 years on the dangers of "data-mining" financial data, particularly deriving general conclusions based on a small amount of information, such as the only U.S. market realization we've had in the past century. There are many instances of statistically-derived trading strategies, attempting to take advantage of market inefficiencies, that fail in future uses. Certainly it is very easy to fall into the trap of overinterpreting financial data, what we generally refer to as overfitting. Avoiding overfitting requires a sound problem formulation, including separate training/testing data sets, care about future information leaking into the model inputs, and out-of-time validation, as well as a good qualitative understanding of the underlying dynamics of the system to be modeled. This understanding of the underlying dynamics guides to (1) avoidance of spurious exogenous variables, (2) appropriate statistical model structures and designs, (3) creation of special model input variables, and (4) a feeling of the robustness of the resulting statistical model.

Misuses of data analysis in financial modeling have been so prevalent that many experts discount any data analysis that is not driven or even dominated by theory. See for example, Fischer Black (1995), who writes "As I have done more theoretical work than empirical work, you may have anticipated my conclusion (about data mining): I find theory to be far more powerful than data when we're trying to estimate expected return. When I read an empirical paper, I usually seek out the theory section and ignore the tables." Probably the major reason that statistical modeling has so often been misused in financial analysis is that the potential benefits are so large that many people are drawn in to this data approach, even with little or no experience in the practice. On the other hand, there's no denying that statistical modeling has been and will continue to provide successful trading opportunities.

The above statement by Black may represent one extreme of the differing points of view, and another might be using a finite data set (the only kind that exist for reality) to extrapolate predictions based on past detailed effects that are not statistically significant. An example of the latter might be to believe that the kinks in the measured distributions in Figure 13 are significant, and that they will reoccur in the future. On the other hand, what should we believe regarding the striking deviation of the general shape, in particular the different sigmas that we see in this figure as holding time increase? Should we listen to the evidence that mean reversion exits or should we hold fast to our standard theory? Many researchers take the middle road and propose modifications of stochastic models that include mean reversion forces, with the properties of these forces derived from the data measurements. This point of view leads us down the path of reasoning that we let data guide the modifications of our theory, which is completely appropriate. Where does this path lead?

In the classical RW with lognormal prices, how do we get the critical statistical parameters R and sigma? From finite-length data sets. Over which time periods do we calculate these statistical parameters? How do we extrapolate these to long-term behavior? The ultimate question is what is the most robust way to estimate what is likely to happen in the future.

We all agree to let data guide us to modify theory for short-term returns, particularly the fat tail effects that are so important to quantifying risk and pricing derivatives. There are still many open important questions about the nature of the tails in these short-term returns, but we all agree that short-term returns do indeed deviate from normally distributed.

What then shall we agree about the nature of long-term returns? Evidence clearly shows that some sort of modification needs to be done with respect to the sigma of long-term distributions. Simply projecting the sigma using classical theory is very different from what we observe in reality (Figures 10 - 13, 16 - 18). Adding mean-reverting dynamics to our stock models is certainly the right direction. How should we add these terms and with what strengths? What should we do for bonds? We answer these questions by examining the data and including new terms with characteristics and parameters estimated from the data. Our theory continues to move more closely to the data, which is appropriate.

Our point of view is that using empirically-derived distributions, with appropriate smoothing, is a closer representation of likely future behavior than a simple model, and usually even better than a complex model. This data-driven approach is successfully used in many other fields, such as engineering, control of complex processes, statistical modeling of consumer behavior, corporate credit risk modeling, etc. Even in these areas of successful application there are many examples of misuse, primarily from overfitting (a common problem with neural nets) or in poor problem design. With appropriate caution, it's hard to model future behavior through theory better than what the past data has directly said. We must never forget that the processes are stochastic and we can only form probability distributions from the data.

Care must particularly be used in examining the tail behavior, which is critical in all risk analysis of markets. We must remain somewhat skeptical about the downside tails in our empirical distributions, although these tails do become more robust in our data as we move to multiple-year time horizons. However, as we move to time horizon windows that have only a few nonoverlapping windows (say 25 years), we again must exercise caution and skepticism. On the other hand, when we compare two instruments as we do in Figure 20, there’s no reason not to have some confidence in the *relative* behavior of the tails. Future stock returns could certainly fall below the stock tail we’ve seen in the past, but the same could be said of bonds, so our relative risk analysis is somewhat more robust.

There are two primary reasons why we hold to theory: (1) it is the most practical way to do analytical treatments of market evolutions and leads to beautiful mathematical representations and insights, and (2) we believe that maintaining a theoretical basis is the safest way to avoid overfitting, or “data-mining”. Both of these sound justifications continue to shrink as our models become more complex and are based on ever-increasing data investigations.

9. Conclusions

In this paper we first described the mainstream approaches to asset allocation. We then demonstrated a robust empirical approach to analyzing long-term market behavior, including the use of overlapping windows and directly and accurately calculating return distributions for long-term holdings. We then discussed a methodology of risk analysis using concepts of Shortfall Risk instead of uncertainty, and discussed how Shortfall Risk is more relevant for long-term analysis. We identify that the choice of methodology for asset allocation calculations is critical, and we present an empirical approach that leads to substantially different results from the standard approaches. We believe this approach and the related results are more realistic and practical.

In examining the assumptions, sensitivities and implications of the mainstream approaches to asset allocation, we find substantial issues. MVO suffers primarily from the fallacy of random walk theory that sigma scales as square root time. The results of MVO depend completely on this assumption, which is very far from true as time horizons grow.

The utility function approach is based on a questionable assumption of the existence of a global UF and in simple cases leads to, at the least, nonuseful results. If the approach doesn't make sense in the simplest case, how can we trust its results in more complex uses? Further, the UF approach has the same drawback as the MVO approach in that mean reversion is not accounted for. Finally, neither of these approaches leads to an explicit quantitative relationship between asset allocation and risk aversion.

There is a danger of heeding the siren song of beautiful mathematics built around questionable assumptions. One can obtain detailed conclusions from long, elegant analysis, simulations and optimizations, and have these results displayed in captivating graphical displays. If however the underlying models are invalid, these beautiful results are simply wrong.

We describe a different approach to analysis of long-term market behavior and use this approach to demonstrate an empirical analysis to quantitative asset allocation. This empirical approach has major benefits over the standard approaches. The empirical approach described here

- Requires no *a priori* assumptions about distributions or market evolutionary dynamics,
- Captures all the appropriate long-term dynamics exhibited by reality, including tail behavior and mean reversion/aversion,
- Allows a direct quantification of risk aversion to dynamic asset allocation optimization strategies,
- Allows the simulation of optimal dynamic asset allocation strategies in the presence of arbitrary complex constraints or consumptions needs.

Using these concepts we measured and compared the relative benefits of long-term holdings in large company stocks to investing in large company bonds. We find the unusual results that

- While stocks exhibit mean reversion for long terms, bonds show the opposite - mean aversion.
- Stock return distributions show *skinny* tails for long-term holdings.
- Bonds are riskier than stocks for holding periods of about 15 years or greater, primarily due to mean reversion in equity markets.

10. Acknowledgement

The authors would like to thank Martin Leibowitz at Morgan Stanley for very useful discussions and comments**.**

Appendix A: Single-Period and Multi-Period Asset Allocation

In this appendix we summarize the single-period mean-variance optimization and the two methods of multi-period asset allocation. The first method of the multi-period allocation is based on the utility function. The optimal asset allocation is determined by optimizing the utility of the consumption over time and/or terminal wealth. The second approach of the multi-period allocation is the direct extension of the single period mean-variance formulation. The optimal asset allocation is determined by the maximizing the return and minimizing the risk at the end of the multi-period.

A.1 Single-Period Mean-Variance Optimization (MVO)

A.1.1 One Risky Asset and One Risk-free Asset

Let the return, standard deviation, and allocation be

risk-free asset: $r_f, \sigma_f = 0, 1-x$
risky asset: $r, \ \sigma, \quad x$

Then the portfolio return and variance are

$$r_p = xr + (1-x)r_f = r_f + x(r - r_f)$$
$$\sigma_p^2 = (x, 1-x)\begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ 1-x \end{pmatrix} = x^2\sigma^2$$

The relationship between return and risk is

$$\sigma_p = \frac{r_p - r_f}{r - r_f}\sigma$$

The efficient frontier is a straight-line.

Let us consider the following optimization problem:

$$\max_x f(x) = r_p - \tfrac{k}{2}\sigma_p^2$$

where k is a risk tolerance factor, when k is smaller (larger), the investor is more aggressive (conservative). Note that

$$f(x) = r_p - \tfrac{k}{2}\sigma_p^2 = r_f + x(r - r_f) - \tfrac{k}{2}x^2\sigma^2 .$$

We have

$$x = \frac{r - r_f}{k\sigma^2} .$$

The Sharpe ratio is:

$$S_p = \frac{r_p - r_f}{\sigma_p} = \frac{r - r_f}{\sigma}.$$

It is the same as the Sharpe ratio of the risky asset.

A.1.2. n Risky Assets and One Risk-free Asset

Let the return, covariance matrix, and allocation be

risk-free asset: $r_f, \sigma_f = 0, 1 - \sum_{i=1}^{n} x_i$

risky assets: $r, \ \Sigma, \qquad x = (x_1, x_2, ..., x_n)^T$.

Then the portfolio return and variance are

$$r_p = \sum_{i=1}^{n} x_i r_i + (1 - \sum_{i=1}^{n} x_i) r_f = r_f + (r - r_f 1)^T x,$$
$$\sigma_p^2 = x^T \Sigma x.$$

Let us consider the optimization problem

$$\max_x f(x) = r_p - \tfrac{k}{2} \sigma_p^2.$$

Note that

$$f(x) = r_p - \tfrac{k}{2} \sigma_p^2 = r_f + (r - r_f 1)^T x - \tfrac{k}{2} x^T \Sigma x$$

and

$$\nabla f(x) = (r - r_f 1) - k \Sigma x.$$

We have

$$x = \tfrac{1}{k} \Sigma^{-1} (r - r_f 1).$$

It can be considered a mutual fund of risky assets. The Sharpe ratio is

$$S_p = \frac{r_p - r_f}{\sigma_p} = \frac{(r - r_f 1)^T x}{\sqrt{x^T \Sigma x}} = \sqrt{(r - r_f 1)^T \Sigma^{-1} (r - r_f 1)}.$$

A.1.3. n Risky Assets

Let the return, covariance matrix, and allocation be: $r, \Sigma, x = (x_1, x_2, ..., x_n)^T$ with a constraint $\sum_{i=1}^{n} x_i = 1$. The return and variance are

$$r_p = r^T x$$
$$\sigma_p^2 = x^T \Sigma x .$$

Let us consider the optimization problem

$$\max_x f(x) = r_p - \tfrac{k}{2}\sigma_p^2$$
$$s.t. \quad 1^T x = 1$$

Introducing a Lagrange multiplier, we have the following Lagrangian:

$$L = r_p - \tfrac{k}{2}\sigma_p^2 - \lambda(1^T x - 1) .$$

Note that

$$\nabla L = r - k\,\Sigma x - \lambda \cdot 1 .$$

We have:

$$x = \frac{1}{k}\Sigma^{-1}\left(r - \lambda \cdot 1\right).$$

Using the constraint, we have

$$\lambda = \frac{1^T \Sigma^{-1} r - k}{1^T \Sigma^{-1} 1}$$

and

$$\begin{aligned} x &= \frac{1}{k}\Sigma^{-1}\left(r - \frac{1^T \Sigma^{-1} r - k}{1^T \Sigma^{-1} 1} \cdot 1 \right) \\ &= \frac{1}{k}\Sigma^{-1}\left(r - \frac{1^T \Sigma^{-1} r}{1^T \Sigma^{-1} 1} \cdot 1 \right) + \Sigma^{-1}\left(\frac{1}{1^T \Sigma^{-1} 1} \right) \end{aligned}$$

It is a combination of 2 mutual funds and the relative weight depends on the risk tolerance k.

Let us consider an example of a portfolio mixed from 2 risky assets. The return and covariance are

$$r = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}, \Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}.$$

Then the asset allocation is

$$x = \frac{1}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \begin{bmatrix} \tfrac{1}{k}(r_1 - r_2) + (\sigma_2^2 - \sigma_{12}) \\ -\tfrac{1}{k}(r_1 - r_2) + (\sigma_1^2 - \sigma_{12}) \end{bmatrix}$$

In the above, we use the Lagrange multiplier method to solve the optimization problem. Alternatively, we can include the constraint directly in the formulation. We use the first asset (0-indexed) from n+1 assets as benchmark.

Asset	*Allocation*
0	$1-\sum_{i=1}^{n} x_i$
1	x_1
...	...
n	x_n

Then the return and variance are

$$r_p = r_0(1-\sum_{i=1}^{n} x_i) + \sum_{i=1}^{n} r_i x_i = r_0 + \sum_{i=1}^{n} (r_i - r_0) x_i$$

$$\sigma_p^2 = (1-\sum_{i=1}^{n} x_i, x_1, ..., x_n)\Sigma \begin{pmatrix} 1-\sum_{i=1}^{n} x_i \\ x_1 \\ ... \\ x_n \end{pmatrix} = x^T Ax + 2x^T B + \Sigma_{00}$$

where

$$A_{ij} = \Sigma_{ij} - \Sigma_{i0} - \Sigma_{j0} + \Sigma_{00}$$
$$B_i = \Sigma_{i0} - \Sigma_{00}.$$

Note that

$$Cov(r_i - r_0, r_j - r_0) = \Sigma_{ij} - \Sigma_{i0} - \Sigma_{j0} + \Sigma_{00},$$
$$Cov(r_i - r_0, r_0) = \Sigma_{i0} - \Sigma_{00}.$$

A and B can be interpreted as the conditional covariance:

$$A_{ij} = Cov(r_i - r_0, r_j - r_0)$$
$$B_i = Cov(r_i - r_0, r_0).$$

Consider the following optimization problem:

$$\max_x f(x) = r_p - \tfrac{k}{2}\sigma_p^2$$

Note that

$$f(x) = r_p - \tfrac{k}{2}\sigma_p^2 = r_0 + (r - r_0 1)^T x - \tfrac{k}{2}(x^T Ax + 2x^T B + \Sigma_{00}).$$

We have

$$\nabla f(x) = (r - r_0) - k(Ax + B).$$

Then we have

$$x = \tfrac{1}{k} A^{-1}(r - r_0 1) - A^{-1}B .$$

This is a combination of 2 mutual funds and the relative weight depends on the risk tolerance factor k.

If we consider the above example (the portfolio from 2 risky assets) in this framework, it is easy to get the same result.

By comparing the results from two different approaches, we have following identities:

$$\Sigma^{-1}\left(r - \frac{1^T \Sigma^{-1} r - k}{1^T \Sigma^{-1} 1} \cdot 1 \right) = A^{-1}(r - r_0 1)$$

and

$$\Sigma^{-1}\left(\frac{1}{1^T \Sigma^{-1} 1} \right) = -A^{-1}B .$$

These matrix identities are very useful when evaluating matrix expressions since sometimes it is easier to evaluate one as opposed to the other.

If we modify the objective function as the linear combination of return and standard deviation, namely

$$\max_x f(x) = r_p - \tfrac{k}{2}\sigma_p .$$

Note that

$$\nabla f(x) = (r - r_0) - \frac{k}{2}\left(\frac{Ax + B}{\sigma_p} \right).$$

We have

$$x = \frac{2\sigma_p}{k} A^{-1}(r - r_0 1) - A^{-1}B .$$

Using the two different objective functions, we get similar results.

A.2. Utility Based Multi-Period Asset Allocation

A.2.1. Hamilton-Jacobi-Bellman (HJB) Equation

The indirect utility function is defined by

$$J(W, x, t) = \sup_{(c_s, \pi_s)_{s \in [t,T]}} E_{W,x,t}\left[\int_t^T e^{-\delta(s-t)} u(c_s) ds + e^{-\delta(T-t)} \overline{u}(W_T) \right].$$

$E_{W,x,t}[.]$ denotes the expectation given that $W_t = W$ and $x_t = x$. c_s is the consumption rate and π_s is the asset allocation at time s. $u(c_s)$ is the utility function of consumption at time s and $\overline{u}(W_T)$ is the utility function at terminal time T.

The indirect utility function in discrete-time form is

$$J_{t_i} = \sup_{(c_{t_n}, \pi_{t_n})_{n=i}^{N-1}} E_{t_i}\left[\sum_{n=i}^{N-1} e^{-\delta(t_n - t_i)} u(c_n)\Delta t + e^{-\delta(T-t_i)}\overline{u}(W_T)\right],\ i = 0,1,...,N-1$$

where $t_i = t_0 + i\frac{T}{N}$. The idea of dynamic programming is that after we make the decision for the first steps, the remaining steps must be optimal in order to reach the global optimal.

$$\begin{aligned}
J_{t_i} &= \sup_{(c_{t_n}, \pi_{t_n})_{n=i}^{N-1}} E_{t_i}\left[\sum_{n=i}^{N-1} e^{-\delta(t_n - t_i)} u(c_n)\Delta t + e^{-\delta(T-t_i)}\overline{u}(W_T)\right] \\
&= \sup_{(c_{t_n}, \pi_{t_n})_{n=i}^{N-1}} E_{t_i}\left[u(c_i)\Delta t + E_{t_{i+1}}\left[\sum_{n=i+1}^{N-1} e^{-\delta(t_n - t_i)} u(c_n)\Delta t + e^{-\delta(T-t_i)}\overline{u}(W_T)\right]\right] \\
&= \sup_{(c_{t_n}, \pi_{t_n})_{n=i}^{N-1}} E_{t_i}\left[u(c_i)\Delta t + e^{-\delta\Delta t} E_{t_{i+1}}\left[\sum_{n=i+1}^{N-1} e^{-\delta(t_n - t_{i+1})} u(c_n)\Delta t + e^{-\delta(T-t_{i+1})}\overline{u}(W_T)\right]\right] \\
&= \sup_{(c_{t_i}, \pi_{t_i})} E_{t_i}\left[u(c_i)\Delta t + e^{-\delta\Delta t} \sup_{(c_{t_n}, \pi_{t_n})_{n=i+1}^{N-1}} E_{t_{i+1}}\left[\sum_{n=i+1}^{N-1} e^{-\delta(t_n - t_{i+1})} u(c_n)\Delta t + e^{-\delta(T-t_{i+1})}\overline{u}(W_T)\right]\right] \\
&= \sup_{(c_{t_i}, \pi_{t_i})} E_{t_i}\left[u(c_i)\Delta t + e^{-\delta\Delta t} J_{t_{i+1}}\right]
\end{aligned}$$

In the fourth equality, the principle of dynamic programming has been used. We reach the Bellman equation:

$$J_{t_i} = \sup_{(c_{t_i}, \pi_{t_i})}\left[u(c_i)\Delta t + e^{-\delta\Delta t} E_{t_i}[J_{t_{i+1}}]\right].$$

This dynamic programming problem is solved by a backward method. First, we make the decision at the last step given the status at that stage. Then we use this recursive Bellman equation to solve the second last step, and so on.

In continuous-time framework, the Bellman equation is

$$J_t = \sup_{(c_t, \pi_t)}\left[u(c_t)\Delta t + e^{-\delta\Delta t} E_t[J_{t+\Delta t}]\right].$$

Rearranging the Bellman equation, we have:

$$\frac{e^{\delta\Delta t} - 1}{\Delta t} J_t = \sup_{(c_t, \pi_t)}\left[e^{\delta\Delta t} u(c_t) + E_t\left[\frac{J_{t+\Delta t} - J_t}{\Delta t}\right]\right].$$

Taking a limit, $\Delta t \to 0$, we have

$$\delta J(W, x, t) = \sup_{(c_t, \pi_t)}\left[u(c_t) + E_t \frac{d}{dt} J(W, x, t)\right].$$

At this time, we must consider the dynamics of wealth. We use the Ito's lemma to include the dynamics of wealth into this equation to reach the Hamilton-Jacobi-Bellman equation.

Let us consider one risk-free asset and d risky assets with following returns and covariance:

wealth return covariance

d risky assets: θ_{it}, μ_{it}, σ_{ijt} $i, j = 1, \dots, d$
1 risk-free asset: θ_{0t}, r_t

The stochastic diffusion processes of the wealth are

$$d\theta_{it} = \theta_{it}(\mu_{it}dt + \sum_j \sigma_{ijt} dz_{jt})$$
$$d\theta_{0t} = \theta_{0t}\, r_t dt$$

where dz_{jt} $(j = 1, 2, \dots, d)$ are the independent Brownian motion. Therefore the wealth dynamics is

$$\begin{aligned} dW_t &= d\theta_{0t} + \sum_i d\theta_{it} + (y_t - c_t)\, dt \\ &= \left[W_t\, r_t + \theta_t^T(\mu_t - r_t 1) + y_t - c_t\right]dt + \theta_t^T \sigma_t\, dz_t \\ &= W_t\left[r_t + \pi_t^T \sigma_t \lambda_t\right]dt + (y_t - c_t)\, dt + W_t \pi_t^T \sigma_t\, dz_t \end{aligned}$$

where $\pi_t = \theta_t / W_t$ is the percent of wealth in risky assets and $\lambda_t = \sigma_t^{-1}(\mu_t - r_t 1)$. c_s is the consumption rate and y_s is the labor income at time s. Suppose x_t is a k-dimensional vector of state variables with

$$dx_t = m(x_t)dt + v^T(x_t)dz_t + \hat{v}(x_t)d\hat{z}_t$$

where $v^T(x_t)$ is a k x d matrix and $\hat{v}(x_t)$ is a k x k matrix.

Using Ito's lemma, we have

$$\begin{aligned} dJ = &\frac{\partial J}{\partial t}dt + J_W\left[W_t\left[r_t + \pi_t^T\sigma_t\lambda_t\right] + y_t - c_t\right]dt + J_x^T m(x_t)\, dt + \frac{1}{2}J_{WW}W_t^2\pi_t^T\sigma_t\sigma_t^T\pi_t\, dt \\ &+ \frac{1}{2}Tr\left[J_{xx}(v_t^T v_t + \hat{v}_t\hat{v}_t^T)\right]dt + J_{Wx}W_t\pi_t^T\sigma_t v_t\, dt + \{\dots\}\, dz \end{aligned}$$

Therefore, we have the Hamilton-Jacobi-Bellman (HJB) equation

$$\delta\, J(W, x, t) = \sup_{(c_t, \pi_t)} \begin{bmatrix} u(c_t) + \frac{\partial J}{\partial t} + J_W\left[W_t\left[r_t + \pi_t^T\sigma_t\lambda_t\right] + y_t - c_t\right] + J_x^T m(x_t) \\ + \frac{1}{2}J_{WW}W_t^2\pi_t^T\sigma_t\sigma_t^T\pi_t + \frac{1}{2}Tr\left[J_{xx}(v_t^T v_t + \hat{v}_t\hat{v}_t^T)\right] + J_{Wx}W_t\pi_t^T\sigma_t v_t \end{bmatrix}$$

The terminal condition is $J(W, x, T) = \overline{u}(W_T)$.

In the constant investment opportunity, there is no x in the formulation. Neglecting the labor income y_s, HJB equation becomes

$$\delta\, J(W, x, t) = \sup_{(c_t, \pi_t)}\left[u(c_t) + \frac{\partial J}{\partial t} + J_W\left[W_t\left[r_t + \pi_t^T\sigma_t\lambda_t\right] - c_t\right] + \frac{1}{2}J_{WW}W_t^2\pi_t^T\sigma_t\sigma_t^T\pi_t\right].$$

A.2.2. Asset Allocation with Constant Investment Opportunity

A.2.2.1. General Case: Intermediate Consumption and Terminal Wealth

If r, μ, σ are constants, in order to optimize the objective function with respect to (c, π) we have $\partial J / \partial c = 0$ and $\partial J / \partial \pi = 0$. They lead to

$$u'(c_t) = J_W(W, t)$$

and

$$J_W \, \sigma \, \lambda + J_{WW} \, W_t (\sigma \sigma^T) \pi_t = 0 \, .$$

Therefore, we have the asset allocations in risky assets:

$$\begin{aligned}
\pi_t &= \left[-\frac{J_W}{W \, J_{WW}} \right] (\sigma \sigma^T)^{-1} \sigma \lambda \\
&= \left[-\frac{J_W}{W \, J_{WW}} \right] (\sigma \sigma^T)^{-1} (\mu - r1) \\
&= \left[-\frac{J_W}{W \, J_{WW}} \right] (\sigma^T)^{-1} \lambda
\end{aligned}$$

Let I_u be the inverse of $u'(c)$, then $c = I_u(J_W(W,t))$. The HJB equation becomes

$$\delta \, J(W, x, t) = u(I(J_W(W,t))) + \frac{\partial J}{\partial t}(W,t) - J_W I(J_W(W,t)) + rJ_W W - \frac{1}{2} \lambda^T \lambda \frac{J_W^2}{J_{WW}} \, .$$

From this equation, we can solve the indirect utility function.

Since in general J is a function of wealth W and time t, the total percent of assets in risky assets is changing over time, but the ratio (π_i / π_j) is constant over time.

Since the linearity of wealth dynamics on W, we can argue that if (c_t, π_t) is the solution of the problem with wealth W_t, then (kc_t, π_t) is the solution of the problem with wealth kW_t.

If we use the constant relative risk aversion (CRRA) utility function for both intermediate consumption and terminal wealth,

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma} \text{ and } \overline{u}(W) = \frac{W^{1-\gamma}}{1-\gamma} \, .$$

We have

$$J(kW,t)=\sup_{(c_s,\pi_s)_{s\in[t,T]}} E_{W,x,t}\left[\int_t^T e^{-\delta(s-t)}\frac{(kc_s)^{1-\gamma}}{1-\gamma}ds+e^{-\delta(T-t)}\frac{(kW_T)^{1-\gamma}}{1-\gamma}\right]$$
$$=k^{1-\gamma}J(W,t)$$

Therefore, we have

$$J(W,t)=W^{1-\gamma}J(1,t).$$

Thus, the relative risk aversion (RRA) of the indirect utility is a constant:

$$RRA=-\frac{W\,J_{WW}}{J_W}=\gamma.$$

In this case, the allocation in risky assets and the risk-free asset is constant over time.

A.2.2.2. Special Case: No Intermediate Consumption

Let us consider the special case where there is no intermediate consumption. The indirect utility only comes from the terminal wealth. HJB equation becomes

$$\delta J(W,x,t)=\sup_{\pi_t}\left[\frac{\partial J}{\partial t}(W,t)+J_W\left[W_t\left[r_t+\pi_t^T\sigma_t\lambda_t\right]\right]+\frac{1}{2}J_{WW}W_t^2\pi_t^T\sigma_t\sigma_t^T\pi_t\right].$$

Setting $\partial J/\partial\pi=0$ leads to

$$J_W\,\sigma\,\lambda+J_{WW}\,W_t(\sigma\sigma^T)\pi_t=0.$$

This is the same equation as in the case with consumption:

$$\pi_t=\left[-\frac{J_W}{W\,J_{WW}}\right](\sigma^T)^{-1}\lambda.$$

Then HJB equation becomes

$$\delta J(W,x,t)=\frac{\partial J}{\partial t}(W,t)+rJ_W W-\frac{1}{2}\lambda^T\lambda\frac{J_W^2}{J_{WW}}.$$

Let us consider an example of bonds and stocks with a risk-free asset. The covariance matrix of bonds and stocks and its inversion are

$$\sigma\sigma^T=\begin{bmatrix}\sigma_{BB} & \sigma_{BS}\\ \sigma_{BS} & \sigma_{SS}\end{bmatrix}\text{ and }(\sigma\sigma^T)^{-1}=\frac{1}{\sigma_{BB}\sigma_{SS}-\sigma_{BS}^2}\begin{bmatrix}\sigma_{SS} & -\sigma_{BS}\\ -\sigma_{BS} & \sigma_{BB}\end{bmatrix}.$$

We have

$$(\sigma\sigma^T)^{-1}(\mu - r1) = \frac{1}{\sigma_{BB}\sigma_{SS} - \sigma_{BS}^2}\begin{bmatrix} \sigma_{SS}(\mu_B - r) - \sigma_{BS}(\mu_S - r) \\ \sigma_{BB}(\mu_S - r) - \sigma_{BS}(\mu_B - r) \end{bmatrix}.$$

Therefore the ratio of allocation in bonds and stocks is

$$\frac{\pi_B}{\pi_S} = \frac{\sigma_{SS}(\mu_B - r) - \sigma_{BS}(\mu_S - r)}{\sigma_{BB}(\mu_S - r) - \sigma_{BS}(\mu_B - r)},$$

again, constant over time.

Here are the three special cases:

If $\rho_{BS} = 1$ then $\frac{\pi_B}{\pi_S} = -\frac{\sigma_S}{\sigma_B}$.

If $\rho_{BS} = -1$ then $\frac{\pi_B}{\pi_S} = \frac{\sigma_S}{\sigma_B}$.

If $\rho_{BS} = 0$ then $\frac{\pi_B}{\pi_S} = \frac{(\mu_B - r)/\sigma_B^2}{(\mu_S - r)/\sigma_S^2}$.

A numerical example: $\mu_B = 2.1\%, \sigma_B = 10\%, \mu_S = 8.7\%, \sigma_S = 20.2\%, \rho_{BS} = 0.2, r = 1\%$, we have:

$$\begin{bmatrix} \pi_B \\ \pi_S \end{bmatrix} = \begin{bmatrix} 16\% \\ 84\% \end{bmatrix} \cdot (\pi_B + \pi_S).$$

A.2.2.3. Special Case: No Intermediate Consumption and No Risk-free Asset

Let us consider the case with no intermediate consumption and no risk-free asset. In this case, the wealth dynamics is

$$dW_t = W_t \pi_t^T \mu \, dt + W_t \pi_t^T \sigma \, dz$$
$$\pi_t^T 1 = 1.$$

The HJB equation is

$$\delta J(W, x, t) = \sup_{\pi_t}\left[\frac{\partial J}{\partial t}(W, t) + J_W \left[W_t \pi_t^T \mu\right] + \frac{1}{2} J_{WW} W_t^2 \pi_t^T \sigma\sigma^T \pi_t\right].$$

Because of the constraint, there are only d-1 independent variables. Specifically, the asset allocation is

$$\pi^T = (\pi_1, \pi_2, ..., 1 - \pi_1 - ... - \pi_{d-1}).$$

$\partial J / \partial \pi_i = 0$ leads to

$$J_W(\mu_i - \mu_d) + J_{WW} W_t \left[(\sigma\sigma^T \pi_t)_i - (\sigma\sigma^T \pi_t)_d\right] = 0, \; i = 1, 2, ..., d-1.$$

Define a $(d-1)\times d$ matrix

$$A=\begin{bmatrix}1 & 0 & 0 & 0 & 0 & -1\\ 0 & 1 & 0 & 0 & 0 & -1\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & 0 & 0 & 1 & -1\end{bmatrix}.$$

We have

$$\frac{J_W}{WJ_{WW}}A\mu + A(\sigma\sigma^T)\pi = 0.$$

Combining with the constraint, we have

$$\begin{bmatrix}A\sigma\sigma^T\\ 1^T\end{bmatrix}\pi = \begin{bmatrix}-\frac{J_W}{WJ_{WW}}A\mu\\ 1\end{bmatrix}.$$

Therefore the allocation is

$$\pi = \begin{bmatrix}A\sigma\sigma^T\\ 1^T\end{bmatrix}^{-1}\begin{bmatrix}-\frac{J_W}{WJ_{WW}}A\mu\\ 1\end{bmatrix}.$$

Let us consider a special case with bonds and stocks only,

$$\sigma\sigma^T = \begin{bmatrix}\sigma_{BB} & \sigma_{BS}\\ \sigma_{BS} & \sigma_{SS}\end{bmatrix} \text{ and } \mu = \begin{bmatrix}\mu_B\\ \mu_S\end{bmatrix}.$$

We have

$$\begin{bmatrix}\pi_B\\ \pi_S\end{bmatrix} = \frac{1}{\sigma_{BB}+\sigma_{SS}-2\sigma_{BS}}\begin{bmatrix}-\frac{J_W}{WJ_{WW}}(\mu_B-\mu_S)+\sigma_{SS}-\sigma_{BS}\\ \frac{J_W}{WJ_{WW}}(\mu_B-\mu_S)+\sigma_{BB}-\sigma_{Bs}\end{bmatrix},$$

again, constant over time.

A.3. Multi-Period Mean-Variance Optimization

A.3.1. Problem Formulation

The multi-period dynamic asset allocation problem is to find the asset allocation, u(t), (t = 0, 1, …, T-1) giving the initial wealth W(0):

$$P(k): \max \ J(u;k) = E[W(T)] - kVar[W(T)]$$
$$= -k\,E[W^2(T)] + k\,E^2[W(T)] + E[W(T)]$$

$\Pi_{P(k)}$ is the solution set.

Because of the term $E^2[x(T)]$, this problem is difficulty to solve using the dynamic programming method.

Let us consider the following similar problem with the parameter λ:

$$P(k,\lambda): \max \ J(u;k,\lambda) = -k\,E[W^2(T)] + \lambda\,E[W(T)]$$
$$= E[-k\,W^2(T) + \lambda\,W(T)]$$

$\Pi_{P(k,\lambda)}$ is the solution set.

The problem $P(k,\lambda)$ can be solved in a dynamic programming framework. We can first solve the problem $P(k,\lambda)$. Then by appropriately choosing λ, one solution is also a solution of P(k). This statement is based on the following result:

If $u_0 \in \Pi_{P(k)}$ then $u_0 \in \Pi_{P(k,\lambda_0)}$, where $\lambda_0 = 1 + 2k\,E[W_0(T)]$.

This can be proved by considering the following function:

$$f(x,y) = -k\,x + k\,y^2 + y, \ where \ \ x = E[W^2(T)], y = E[W(T)].$$

Since the Hessian matrix of f(x, y) is semi-positive definite, we have the following inequality using the Taylor expansion:

$$(-k\,x + k\,y^2 + y) - (-k\,x_0 + k\,y_0^2 + y_0) \geq (-k\,x + \lambda\,y) - (-k\,x_0 + \lambda\,y_0),$$

where $\lambda = 1 + 2k\,y_0$. Suppose that $u_0 = (x_0, y_0)$ is the best in $\Pi_{P(k)}$, but is not the best in $\Pi_{P(k,\lambda)}$, then there exists (x, y) such that:

$$-k\,x + \lambda\,y > -k\,x_0 + \lambda\,y_0.$$

Thus we have

$$-k\,x + k\,y^2 + y > -k\,x_0 + k\,y_0^2 + y_0.$$

Namely, $u_0 = (x_0, y_0)$ is not the best in $\Pi_{P(k)}$. This is a contradiction, therefore the proposition is proved.

A.3.2. Discrete-Time Stochastic Dynamic Programming

Let us consider n+1 assets:

Return wealth

$$0 \quad r_0(t) \quad W(t) - \sum_{i=1}^{n} u_i(t)$$

$$1 \quad r_1(t) \quad u_i(t)$$

......

$$n \quad r_n(t) \quad u_n(t)$$

where t = 0, 1, …, T-1. The wealth dynamics is

$$\begin{aligned} W(t+1) &= (W(t) - \sum_{i=1}^{n} u_i(t)) r_0(t) + \sum_{i=1}^{n} u_i(t) r_i(t) \\ &= W(t) r_0(t) + P^T(t) u(t) \end{aligned}$$

where $P^T(t) = (r_1(t) - r_0(t), ..., r_n(t) - r_0(t))$. The objective function in dynamic programming is

$$\max_{\{u(t), t=0,1,...,T-1\}} E[-W^2(T) + \gamma W(T)],$$

where $\gamma = \lambda / k$ is a parameter that is determined after we solve this problem. We need to make multiple decisions from time 0 to time T-1 to optimize the objective function. The dynamic programming method is suitable for this kind of problems. The problem can be solved by backward approach.

Let us first consider the last step at time T-1. Given the wealth at time T-1, W(T-1), we have the objective function as

$$J(W(T-1)) = \max_{u(T-1)} E[-W^2(T) + \gamma W(T)].$$

Suppose we have made all decisions except the last step, from the optimality principle in dynamic program that the final step must be conditionally optimal.

Note that the wealth dynamics at time T-1 is

$$W(T) = W(T-1) r_0(T-1) + P^T(T-1) u(T-1).$$

We have

$$\begin{aligned} J(W(T-1)) &= \max_{u(T-1)} E[-W^2(T) + \gamma W(T)] \\ &= \max_{u(T-1)} E \begin{bmatrix} -W^2(T-1) r_0^2 - u^T(T-1) pp^T u(T-1) - 2(r_0 p) u(T-1) W(T-1) \\ + \gamma W(T-1) r_0 + \gamma p^T u(T-1) \end{bmatrix} \end{aligned}$$

It is a quadratic form of u(T-1) and it is easy to determine u(T-1):

$$u(T-1) = \frac{\gamma}{2} E^{-1}(PP^T) E(P) - E^{-1}(PP^T) E(r_0 P) W(T-1).$$

Note that all expectations are calculated in time period T-1, for example E(P) = E(P(T-1)). Inserting the results of u(T-1) and W(T) into J(W(T-1)) we have

$$J(W(T-1)) = -W^2(T-1)\,A_2(T-1) + \gamma W(T-1)\,A_1(T-1) + \frac{\gamma^2}{4}E(P^T)E^{-1}(PP^T)E(P)$$

where

$$A_1(t) = E(r_0(t)) - E(P^T(t))E^{-1}(P(t)P^T(t))E(r_0P(t))$$
$$A_2(t) = E(r_0^2(t)) - E(r_0(t)P^T(t))E^{-1}(P(t)P^T(t))E(r_0P(t))$$

Now we consider the step at time T-2. In terms of the principle of optimality, the conditional objective function at T-2 is

$$\begin{aligned}
J(W(T-2)) &= \max_{\substack{u(T-1)\\u(T-2)}} E[-W^2(T) + \gamma W(T)] \\
&= \max_{u(T-2)} E\left[\max_{u(T-1)} E[-W^2(T) + \gamma W(T)]\right] \\
&= \max_{u(T-2)} E(J(W(T-1))) \\
&= \max_{u(T-2)} E\left\{const * \left[-W^2(T-1) + \left(\gamma \frac{A_1(T-1)}{A_2(T-1)}\right)W(T-1)\right] + const\right\} \\
&\sim \max_{u(T-2)} E\left[-W^2(T-1) + \left(\gamma \frac{A_1(T-1)}{A_2(T-1)}\right)W(T-1)\right]
\end{aligned}$$

Note that this form is the same as the objective function at time T-1 by replacing

$$\gamma \to \gamma \frac{A_1(T-1)}{A_2(T-1)}.$$

This recursive form of the conditional objective function is very critical in dynamic programming. Then we can write down the solution for u(T-2) immediately:

$$u(T-2) = \frac{\gamma}{2}\frac{A_1(T-1)}{A_2(T-1)}E^{-1}(PP^T)\,E(P) - E^{-1}(PP^T)\,E(r_0P)W(T-2).$$

In general, we can write this down for t = 0, 1, 2, … T-1 as

$$u(t) = \frac{\gamma}{2}\left(\prod_{k=t+1}^{T-1}\frac{A_1(k)}{A_2(k)}\right)E^{-1}(P(t)P^T(t))\,E(P(t)) - E^{-1}(P(t)P^T(t))\,E(r_0(t)P(t))W(t).$$

The recursive equations for W(t) and $W^2(t)$ are

$$E[W(t+1)] = A_1(t)W(t) + \frac{\gamma}{2}\left(\prod_{k=t+1}^{T-1}\frac{A_1(k)}{A_2(k)}\right)E(P^T(t))E^{-1}(P(t)P^T(t))\,E(P(t))$$
$$E[W^2(t+1)] = A_2(t)W(t) + \frac{\gamma^2}{4}\left[\prod_{k=t+1}^{T-1}\left(\frac{A_1(k)}{A_2(k)}\right)^2\right]E(P^T(t))E^{-1}(P(t)P^T(t))\,E(P(t)).$$

Note that from the recursive equation, $f(t+1) = a(t) f(t) + b(t), (t = 0,1,...,T-1)$, the explicit expression for f(T) is

$$f(T) = \left(\prod_{t=0}^{T-1} a(t) \right) f(0) + \sum_{t=0}^{T-1} \left(\prod_{k=t+1}^{T-1} a(k) \right) b(t).$$

Therefore, the expectations for the terminal value are

$$E[W(T)] = \mu W(0) + \nu \gamma$$
$$E[W^2(T)] = \tau W^2(0) + \frac{\gamma^2}{2} \nu,$$

where

$$\mu = \prod_{t=0}^{T-1} A_1(t)$$
$$\nu = \sum_{t=0}^{T-1} \left(\prod_{k=t+1}^{T-1} A_1(k) \right) \frac{1}{2} \left(\prod_{k=t+1}^{T-1} \frac{A_1(k)}{A_2(k)} \right) E(P^T(t)) E^{-1}(P(t) P^T(t)) E(P(t))$$
$$\tau = \prod_{t=0}^{T-1} A_2(t).$$

Now we have determined all allocations at all periods with the parameter λ. The parameter can be determined by

$$\lambda = k\gamma = 1 + 2k E(W(T)) = 1 + 2k(\mu W(0) + \nu \gamma).$$

We have:

$$\gamma = \frac{2\mu}{1-2\nu} W(0) + \frac{1}{k(1-2\nu)}.$$

Therefore, the mean and variance are

$$E[W(T)] = \mu W(0) + \nu \gamma = W(0) \left(\frac{\mu}{1-2\nu} \right) + \frac{\nu}{k(1-2\nu)}$$
$$Var[W(T)] = \tau W^2(0) + \frac{\gamma^2}{2} \nu - (\mu W(0) + \nu \gamma)^2 = \left(\tau - \frac{\mu^2}{1-2\nu} \right) W^2(0) + \frac{\nu}{2k^2(1-2\nu)}.$$

After eliminating the parameter k, we have the efficient frontier:

$$Var[W(T)] = \left(\tau - \frac{\mu^2}{1-2\nu} \right) W^2(0) + \frac{(1-2\nu)}{2\nu} \left(E(W(T)) - W(0) \frac{\mu}{1-2\nu} \right)^2.$$

In terms of mean return and variance of return the efficient frontier is

$$\sigma_T^2 = \left(\tau - \frac{\mu^2}{1-2\nu} \right) + \frac{(1-2\nu)}{2\nu} \left(r_T - \frac{\mu}{1-2\nu} \right)^2.$$

This is an unconstrained efficient frontier in the sense that some assets could be in a short position.

A.3.3. Asset Allocation with Constant Investment Opportunity

When the return and covariance, (r, Σ), are constants, we can work out the expectations in above formulas. Let

$$a_0 = 1^T \Sigma^{-1} 1$$
$$a_1 = r^T \Sigma^{-1} 1$$
$$a_2 = r^T \Sigma^{-1} r$$

We have

$$A_1 = \frac{a_1}{a_0 + a_0 a_2 - a_1^2}$$
$$A_2 = \frac{1 + a_2}{a_0 + a_0 a_2 - a_1^2}$$
$$E(P^T) E^{-1}(PP^T) E(P) = \frac{a_0 a_2 - a_1^2}{a_0 + a_0 a_2 - a_1^2}$$
$$E^{-1}(PP^T) E(P) = \frac{\Sigma^{-1}\left(a_0\, r - a_1 1\right)}{a_0 + a_0 a_2 - a_1^2}$$ disregard the 1st component on RHS

$$E^{-1}(PP^T) E(r_0 p) = \Sigma^{-1}\left(A_1\, r - A_2 1\right)$$ disregard the 1st component on RHS.

These formulas are useful for calculating the asset allocation in each step and efficient frontier.

Appendix B. After-Inflation, Real Returns Analysis

All of the analysis in the paper has been done before inflation. It is easy and very useful to do the same analysis using after-inflation, or real returns, where we continuously adjust for CPI. We did the major analysis before inflation to keep it simpler, but we believe that the after-inflation analysis would generally be more relevant for investors.

Figure B1 shows the expected returns at different confidence levels for different holding periods for stocks, after inflation is removed. Figure B2 shows the same for large company bonds.

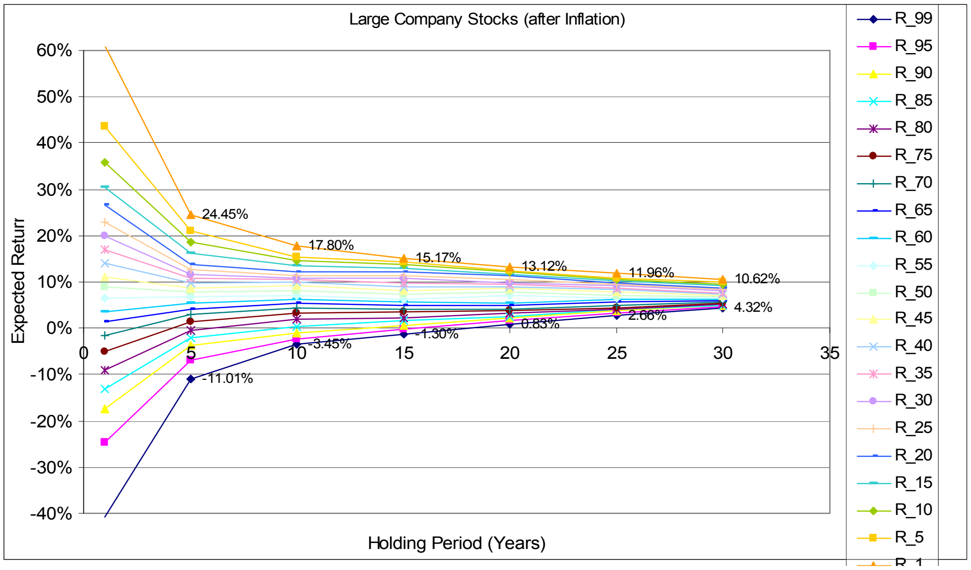


Figure B1. Stocks returns after inflation as a function of holding period.

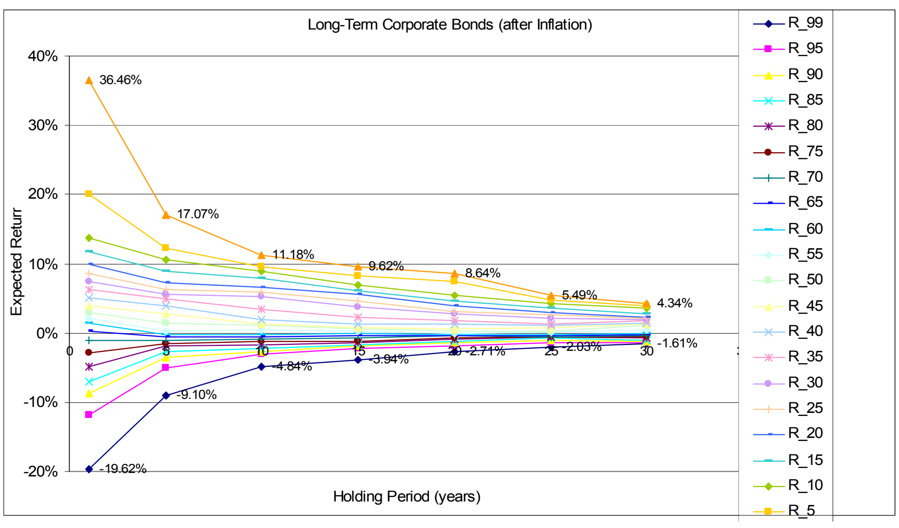


Figure B2. Long-term corporate bond returns after inflation as a function of holding period.

Appendix C. Investigation of Different Data Time Windows

For all time series analysis it's important to carefully question what data one should consider. For example, if one wants to build GARCH volatility models one carefully selects a past time window for training that is believed to be relevant.

Any application and general beliefs derived from financial modeling guided by real data are critically dependent on the data set one selects. This truism is another typical criticism of empirical studies, such as this one, but we submit that any theoretically-derived result has just as critical dependence on the statistically-measured values of R and sigma, which depend as strongly on the choice of data set. Guided by historical data, Figure 15, we might reasonably conclude that there is a fundamental change in market behavior around 1970, particularly for bonds. We might also reasonably decide that a market event as strong as the Great Depression has low likelihood of reoccurring. Based on this reasoning, we present additional studies using only the first and second halves of this data set separately. Figure C1 shows the decay of sigma for stocks using (1) all data (same as previous Figure 12), (2) data from 12/1925 through 12/1964, and (3) data from 12/1964 through 12/2003. When we use first or second half data we pretend that's the only data we have, so our overlapping windows must be contained in these time frames. Because of this we recognize that our confidence in the behavior of the curves declines beyond about a 15-year holding period. Figure C2 shows the same for Bonds.

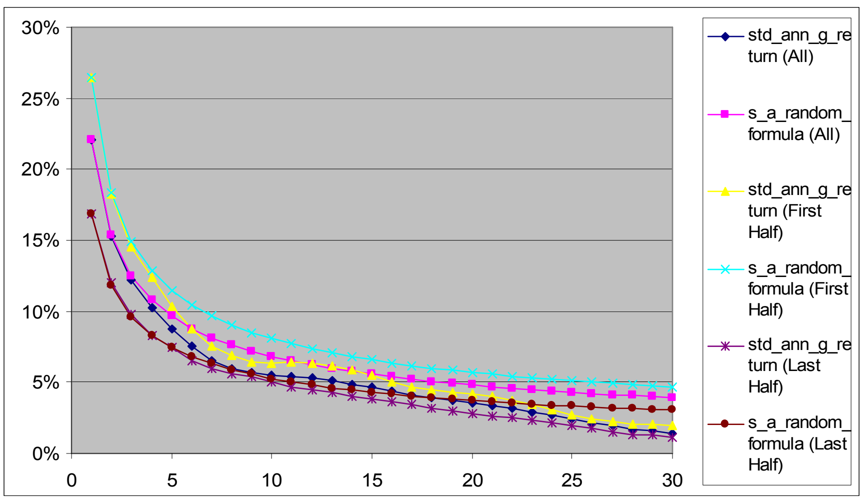


Figure C1. Time decay of sigma for stocks as a function of holding period for (1) all data, (2) data from 1926 through 1964, and (3) data from 1964 through 2003.

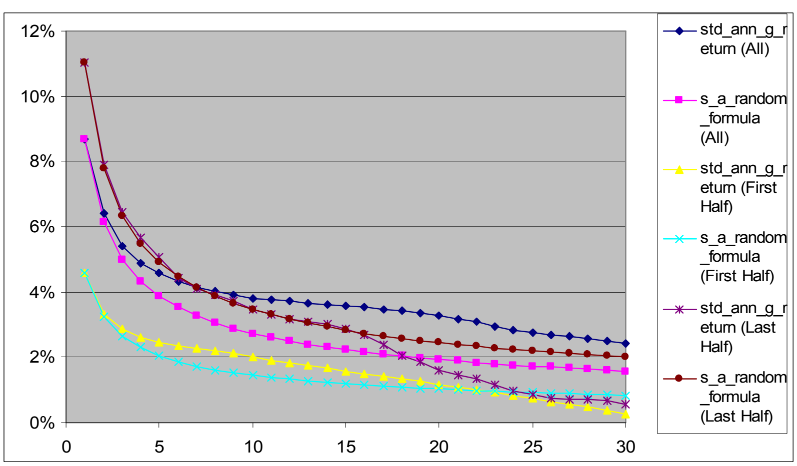


Figure C2. Time decay of sigma for bonds as a function of holding period for (1) all data, (2) data from 1926 through 1964, and (3) data from 1964 through 2003.

In the Stock picture, Figure C1, we see that

- Volatility is generally higher in the first half, followed by the "all", and lowest in the second half of the data.
- All periods show mean reversion - the sigma falls faster than square root time
- The second half shows the least mean reversion

In the Bond picture, Figure C2, we see that

- Volatility is highest in the last half, followed by the "all", and lowest in the first half.
- The first half and the "all" show strong mean aversion, and the last half shows a good match to the classical square root time scaling, with perhaps mean reversion after about 15 years, although our confidence in this is not strong.

The high volatility in stocks in the first half is due primarily to the large "Great Depression" event. It is likely that bonds had a fundamental shift in behavior during and perhaps after the period of hyperinflation in the 70's.

Figures C3 and C4 show a comparison of stocks to bonds for the first half (C3), second half (C4) of the data.

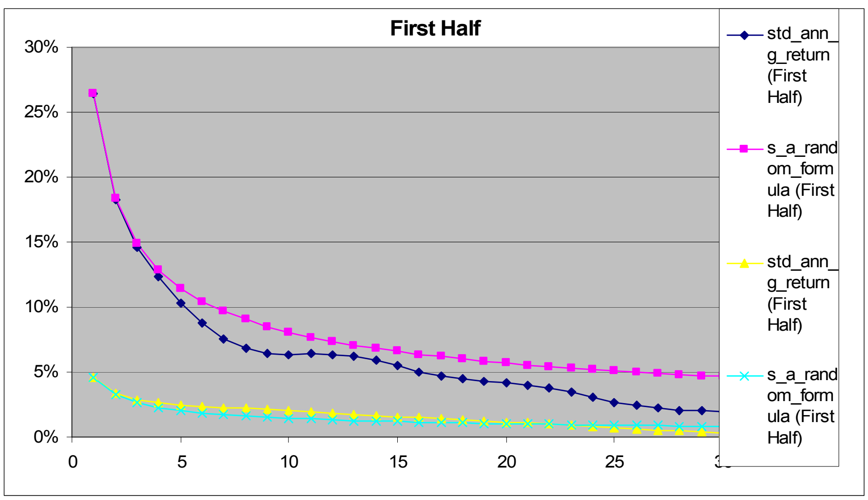


Figure C3. Decay of stocks and bonds sigmas as a function of holding period, theoretical and actual. Data from 1926 through 1964.

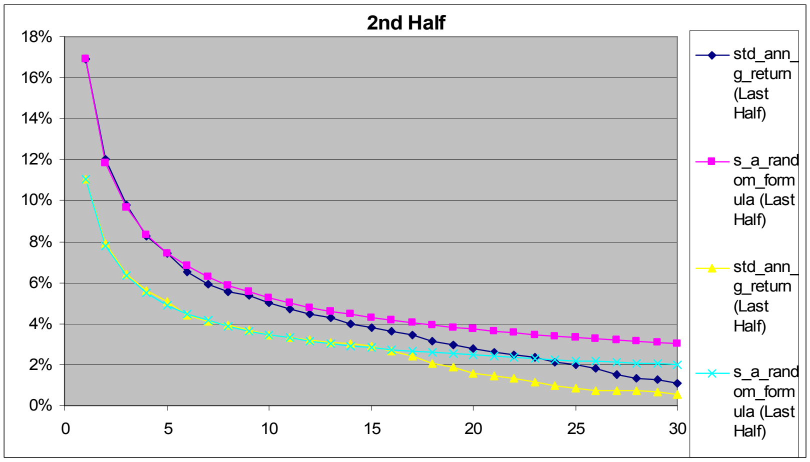


Figure C4. Decay of stocks and bonds sigmas as a function of holding period, theoretical and actual. Data from 1964 through 2003.

Some argue that the best data sets would be "all" for stocks and the "second half" for bonds, shown in Figure C5.

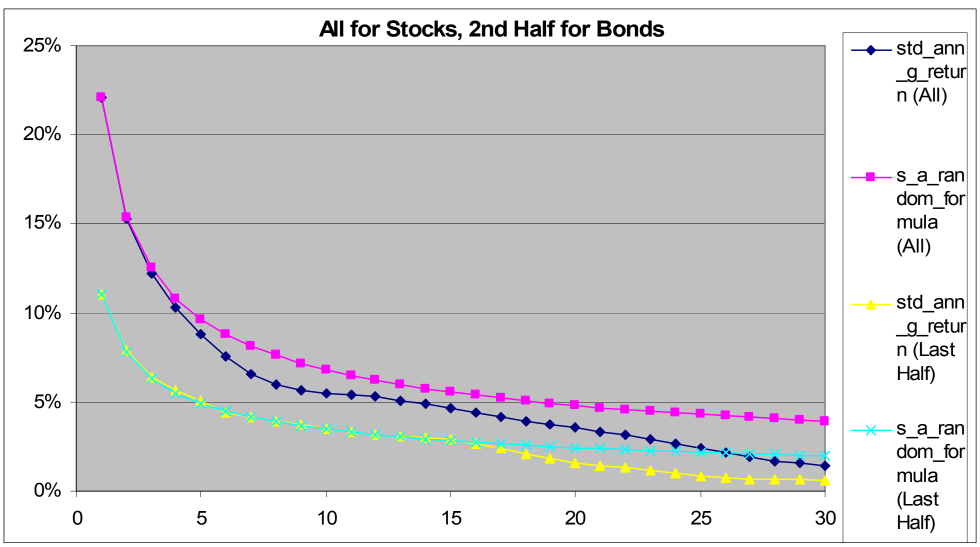


Figure C5. Decay of stocks and bonds sigmas as a function of holding period, theoretical and actual. Stock data from 1926 through 2003, bond data from 1964 through 2003.

Clearly any results and conclusions depend critically on the choice of data sets, regardless of what approach one takes. What we care about is what are the best data sets to study that we believe have the most relevance to what we think is likely to happen in the future. One can argue about whether we should throw out bond data before about 1970, as in Figure B4, but we could just as reasonably argue that the Great Depression won't happen again because there are now so many more controls in the financial world. Some would argue that the bond data BEFORE 1970 is more representative of the future than the period of hyperinflation. We would argue that the entire data sets are the most relevant since they best represent the spectrum of what might occur in the future, particularly in terms of long-time-horizon behavior. Certainly there will be events and regimes that we haven't yet seen, and we would argue that one should use the largest data sets available where the data is valid in order to best approximate the range of future possibilities.

References

J. Siegel, *Stocks for the Long Run*, 3rd Edition, McGrall-Hill, (2002).

Ibbotson Associates, *Stocks, Bonds, Bills, Inflation*, 2004 Yearbook, Ibbotson, (2004).

P. Gopikrishnan et al., “Scaling of the distribution of fluctuations of financial market indices”, Phys. Rev. E, 60, p. 5305 (1999).

J. Doyne Farmer, “Physicists Attempt to Scale the Ivory Towers of Finance”, Computing in Science and Engineering, Nov/Dec 1999, p. 26.

B. Malkiel, *A Random Walk Down Wall Street,* W. W. Norton, (1996).

J. Bogle, *Common Sense on Mutual Funds,* Wiley and Sons, (1999).

Andrew W. Lo and Craig MacKinlay, *A Non-Random Walk Down Wall Street*, Princeton University Press, 2002.

J.M. Poterba and L.H. Summers, “Mean Reversion in Stock Prices: Evidence and Implications”, Journal of Financial Economics 22, page 27 (1988).

M. Richardson and T. Smith, “A Unified Approach to Testing for Serial Correlation in Stock Returns”, Journal of Business 63, page 371 (1994). They argue that overlapping windows are better.

Peter Bernstein, *Against the Gods: The Remarkable Story of Risk,* John Wiley and Sons (1996).

Martin Leibowitz, *Investing*: *The Collected Works of Martin Leibowitz,* Frank Fabozzi editor, Probus (1992).

Fischer Black, “Estimating Expected Return”, Financial Analysts Journal, Jan-Feb (1995).

David Kahneman and Amos Tversky, *Choices, Values and Frames*, Cambrige University Press (2000).

Mattew Rabin and Richard Thaler, “Risk Aversion”, Journal of Economic Perspectives, 15, (2001).

Michael Stutzer, “Asset Allocation Advice: Reconciling Expected Utility and Shortfall Risk”, preprint (2003).

Benoit Mandelbrot and Richard Hudson, *The (Mis)Behavior of Markets*, Basic Books, 2004.