Portfolio and Investment Analysis with SAS®

Financial Modeling Techniques for Optimization

John B. Guerard
Ganlin Xu
Ziwei Wang

The correct bibliographic citation for this manual is as follows: Guerard, John B., Ziwei Wang, and Ganlin Xu. 2019. *Portfolio and Investment Analysis with SAS®: Financial Modeling Techniques for Optimization*. Cary, NC: SAS Institute Inc.

Portfolio and Investment Analysis with SAS®: Financial Modeling Techniques for Optimization

978-1-64295-193-6 (Hard cover)
978-1-63526-692-4 (Hardcopy)
978-1-63526-691-7 (Web PDF)
978-1-63526-689-4 (epub)
978-1-63526-690-0 (mobi)

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

March 2019

Contents

About This Book

What Does This Book Cover?

In this applied investment book, we introduce the risk-return tradeoff of financial investments as well as stocks and bonds investing in the US and global markets over the 2002–2016 time period. Stocks have produced higher rates of return relative to risk in global markets than US markets. We report why intelligent investors prefer more stocks than bonds to maximize stockholder wealth. The bulk of this book is concerned with demonstrating how individuals, whether students or real-world investors, can select stocks and create portfolios to maximize expected returns for a given level of risk. The authors do not believe in completely Efficient Markets, and we show how to outperform the markets by using sophisticated statistical modeling implemented in SAS. The authors believe that “Quant” life is pass/fail. Your models are either statistically significant, or they are not.

The authors have used SAS for over 35 years in financial modeling and investment research. We stress the need to generate statistically significant stock selection modeling and portfolio construction management and measurement. The authors believe that the Markowitz Efficient Frontier can be applied by students, investors, and Certified Financial Planners using SAS to create variables, run robust regression models for stock selection, and use the stock expected returns and risk inputs to create Efficient Frontiers.

Is This Book for You?

We assume no previous knowledge of finance, investments, statistics, or optimization. The text shows you how to analyze income statements, balance sheets, and sources and uses of funds statement analyses. We show why some variables are more useful to consider for model building based on Information Coefficients (ICs) and estimated Efficient Frontiers with realistic transactions costs. The authors discuss stock universes developed and modeled from the perspective of ICs and decile spreads. We report why variables might be different in US, Chinese, Japanese, European, Emerging Market, and global stock universes with respect to a large set of variables of analysts’ earnings per share forecasts, forecast revisions, and direction of revisions. The authors have written technical papers and texts that are included in the References section. In this book, we want to introduce analysis beyond the typical undergraduate investments course to help enhance portfolio returns.

We use PROC REG, PROC IML, and PROC ROBUSTREG to run monthly regressions and to demonstrate how to address outlier issues (Beaton-Tukey and Tukey Optimal Influence Function weighting schemes) and multicollinearity issues. We refer to regression using Beaton-Tukey outlier-adjusted weighting and principal components (PCA) analysis as WLRR analysis. Regression modeling using US stocks is referred to as US Expected Returns (USER) and using global stocks as Global Expected Returns (GLER). Regression modeling will use various (M, S, and MM) robust procedures and various (Huber, Hampel, Andrews, Tukey, and Yohai) weighting schemes. The MM methods, using the Tukey and Yohai Optimal Influence Functions, enhance stock selection modeling.

The authors develop variations on Markowitz and Sharpe portfolio optimization techniques, which will illustrate the relative efficiency of individual variables (sales, earnings, book value, dividends, cash flow, forecasted earnings, EP, BP, DP, SP, CP, and FEP) and robust regression-weighted stock selection models.

The authors develop and test the Markowitz-Xu Data Mining Corrections (DMC) procedure and compare it with more recently developed tests. We report statistically significant DMC results. We have significant experience as teachers and practitioners in financial theory, valuation, and financial modeling. We have intimate knowledge of the data available to bridge the theory and application and show how to enhance portfolio returns and maximize terminal wealth.

What Should You Know about the Examples?

This book includes tutorials for you to follow to gain hands-on experience with SAS.

Software Used to Develop the Book's Content

We use SAS 9.3 TS1M0 for Windows 7 Professional 32-bit system to run all the SAS code.

Example Code and Data

You can access the example code and data for this book by linking to its author page at support.sas.com/guerard.

SAS University Edition

This book is compatible with SAS University Edition. If you are using SAS University Edition, then begin here: https://support.sas.com/ue-data .

We Want to Hear from You

Do you have questions about a SAS Press book that you are reading? Contact us at saspress@sas.com.

SAS Press books are written *by* SAS Users *for* SAS Users. Please visit sas.com/books to sign up to request information on how to become a SAS Press author.

We welcome your participation in the development of new books and your feedback on SAS Press books that you are using. Please visit sas.com/books to sign up to review a book

Learn about new books and exclusive discounts. Sign up for our new books mailing list today at https://support.sas.com/en/books/subscribe-books.html.

Learn more about these authors by visiting their author pages, where you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more:
support.sas.com/guerard
support.sas.com/xu
support.sas.com/zwang

Chapter 1: Why Do We Invest?

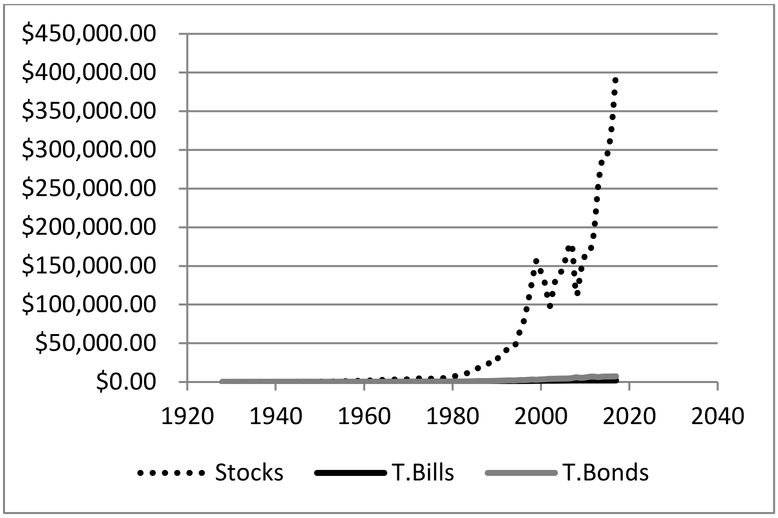
1.1 Introduction

Consumers balance their current needs and wants of consumption – spending on food, housing, and other expenses – with their desires for future consumption, such as educational expenses, vacations, or buying a long-desired sports car at age 65. People save from their current income to invest in assets that will grow over time so that they can consume more in the future. The purpose of this book is to show readers how use SAS to enhance their wealth.

The total US household net wealth has grown from net wealth of $87 trillion at the end of 2016 to $100 trillion, as reported by the US government (Board of Governors of the Federal Reserve System (2018)), see Torry (2018). The stock market's 20-plus percent return in 2017 is a significant contribution to this jump of wealth. Out of the $100 trillion, $28 trillion are earmarked as retirement assets, which have various tax-favored treatments. The Census report shows that at least 75 percent of households had at least $50,000 in net assets at the end of 2011. Those assets in a household's portfolio might consist of cash, bond, stocks, real estate and so on. These different assets provide different return and risk characteristics. The Federal Reserve reported that the median financial asset value is $200,000 for people 40 years older.

Everybody is a financial asset investor, either passively or actively. If you are the owner of a house, you are an investor in real-estate. It is often stated that your real estate investment is the largest investment a family will make. A recent examination of the Case-Shiller Housing Index reveals that the US housing market has reached an all-time high, as measured from 1970. If you have 401k accounts, you are likely a passive investor. If you have a brokerage account and trade a lot, you are active investor. Even the pension amount that you receive at retirement depends on the market performance. Figure 1.1 provides the cumulative wealth of an investor who invested $100 at beginning of 1928 in either cash, bond, or US stock market.

Figure 1.1: Cumulative Wealth of $100 Starting in 1928

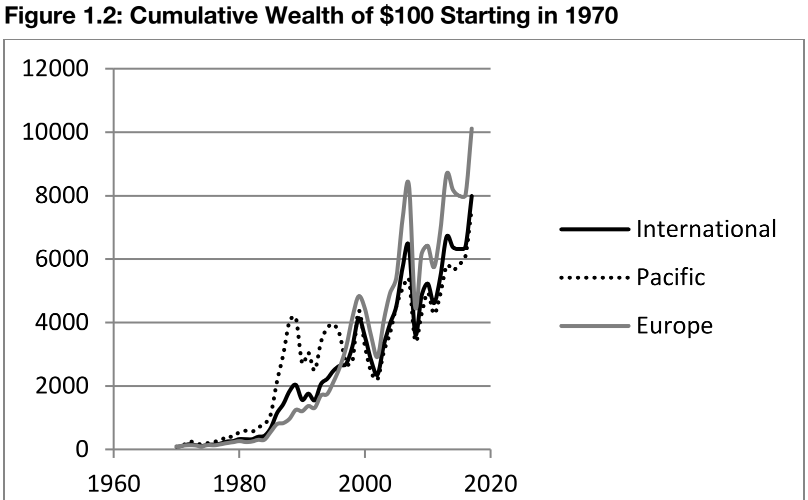


As Figure 1.1 shows, stocks have outperformed bonds, and bonds have outperformed cash. One hundred US dollars at the beginning of 1928 has turned into almost $400,000 at the end of 2017, a return of 400,000 percent for a holding period of 90 years, despite the fact that this period included the Great Depression of 1930 and the 2008 recession. Wealth accumulation does not grow monotonously. The stock market lost about 40 percent of its value recently in year 2001–2002, and year 2008. We use these three types asset as example because they have the longest return history. The S&P/Case-Shiller US National Home Price Index published by the Federal Reserve Bank of St. Louis, a popular index for the real-estate asset class, was started in 1987.

The American people have recognized the earnings power and the risk of stock. The Organization for Economic Co-operation and Development (OCED) reports that only 13.5 percent of US household's financial asset is cash. Xu (2015) reported that young Americans invest more than 80 percent in stocks in their 401k accounts.

The wealth creation power of equity is not limited to the US. Figure 1.2 shows wealth accumulation by investing in international stocks. One hundred dollars invested at the beginning of 1970 in the European market turned into $10,000 at the end of 2017, a return of 10,000 percent with a holding period of 47 years.

Figure 1.2: Cumulative Wealth of $100 Starting in 1970



1.2 Assumptions

All writers have beliefs, even prejudiced views. We will disclose our beliefs at the outset. We believe in Active Quantitative Management using the portfolio selection, construction, and management techniques of Harry Markowitz, William (Bill) Sharpe, Jan Mossin, John Blin, Henry Latane, Martin Gruber and Ed Elton, Barr Rosenberg, Haim Levy, and the investment professionals at Factset, FIS, and Axioma. We believe that the empirical evidence of the past 30 years suggests that financial anomalies were identified, have persisted, and most likely will persist into the coming decade.

We believe the benchmarks established by Markowitz, Sharpe and Blin are still relevant, and difficult to beat. New data, better computational power, and enhanced statistical analyses are shown and discussed in this text. We believe that new data, also known as "Big Data," will enhance returns in the future, but the enhancements will be more in the 15–20% range, rather than doubling existing excess returns. We present models, updated analyses, and evidence to show that robust regression, as we estimated in financial models nearly 30 years ago, still works today.

Earnings forecasting can be used to greatly enhance portfolio returns in US, and particularly, non-US markets. Financial models, when properly developed and tested with proper transactions cost, work about 75 percent of the time. The models produce statistically significant excess returns in the years that the models win, but ONLY if the models are used religiously, 100 percent of the time and asset owners fully invest the Mean-Variance weights (or Equal Active Weights, plus or minus the benchmark, at least two percent active weighting, which generally produce lower Sharpe ratios).

Henry Latane and Harry Markowitz taught us in 1959 that to maximize the Geometric Mean maximizes the utility of final wealth; achieving the greatest level of terminal wealth, in the shortest time possible. Henry Latane, in his UNC Portfolio Analysis doctoral seminar, often joked that the Efficient Markets hypothesis only said that the average investor only earned an average return, adjusted for risk. Latane asked, "Who wants to be average?" We believe that smart people, with good databases, can enhance returns about 1–2 percent, annualized, adjusted for risk and risk premiums accepted (knowing or unknowingly incurred). The authors detest closet benchmark-huggers.

1.3 Annualized Return

Cumulative wealth is the product of multi-period returns. If an asset has returns R_t, for t=1,…,T, where t is the period, and T is the total periods, then cumulative wealth is calculated as

$$W_T = W_{T-1} * (1+R_T) = W_0 \cdot \prod_{t=1}^{T} (1+R_t) = W_0 \cdot (1+R_{HP}) \tag{1}$$

where R_{HP} is the holding period return. Even though the holding period returns are correct numbers to capture the wealth change, it is difficult to compare the merit of asset returns with different holding periods. In our case, we have US return data going back to 1926 and MSCI international data going back to 1970. We need a number for merit to compare the returns of US stock and international stock, one with a 90-year holding period, and one with a 47-year holding period. To facilitate comparison, financial reports often use a one-period return instead of holding period return. This one-period return is called the *annualized return*, which is calculated from the holding period return. If this calculated return is realized for every period, then the cumulative wealth will be equal with the wealth generated by the actual holding period return.

$$W_T = W_0 * \prod_{t=1}^{T} (1+R_t) = W_0 * (1+R_g)^T$$

From this equality, we derive the formula

$$R_g = \sqrt[\frac{1}{T}]{\prod_{t=1}^{T}(1+R_t)} - 1 \tag{2}$$

Table 1.1 shows that it is much easier to compare the annualized return than to compare the holding period returns. For the long period of 1928–2017, US stocks outperformed bonds by 4.77 percent a year. For the shorter period of 1970–2017, US stocks outperformed bonds by 3.36 percentage points a year. For the same period, the US stocks outperformed the Europe stocks slightly by only 30 basis points per year. The European stocks outperformed the Pacific stocks by 65 basis points.

Table 1.1: Annualized Return in Percentage

Holding Period	US Stock	US Bond	International	Pacific	Europe
1928-2017	9.65	4.88			
1970-2017	10.45	7.09	9.55	9.44	10.10

1.4 Average Return

Although annualized return is an accurate yearly number to describe the earning power of each asset, from a portfolio's point of view, it lacks the additive property. In other words, the portfolio's annualized return is not the portfolio's weighted annualized return. It is therefore more convenient to work with the simple average return in portfolio management, which is

$$R_a = \frac{\sum_{t=1}^{T} R_t}{T} \tag{3}$$

The average return is often called the arithmetic mean. The annualized return is often called the geometric mean. Table 1.2 shows the arithmetic means of the same five assets shown in Table 1.1.

Table 1.2: Average Return in Percentage

Holding Period	US Stock	US Bond	International	Pacific	Europe
1928-2017	11.53	5.15			
1970-2017	11.83	7.50	11.74	12.92	12.20

Notice that the arithmetic means reported in Table 1.2 are always higher than the geometric means reported in Table 1.1. The Pacific stock is the best-performing asset according the arithmetic mean, while US stock is the best-performing asset in reality according to Table 1.1. The arithmetic mean ignores the variability of returns whereas geometric mean couples the variability of return too tightly. The geometric mean is roughly the arithmetic mean subtracted by half the variability

$$R_g \approx R_a - 0.5 * \sigma^2 \tag{4}$$

1.5 Expected Return

We can use SAS to calculate the means returns, return variabilities, and return correlations of the assets from Tables 1.1 and 1.2, and report the results in Output 1.1.

Program 1.1: Correlations of Global Market

```
proc corr data = global_assets_annual_returns;
     var USStocks USBonds Pacific European;
run;
/*Data Source: SBBI(2017)*/
```

Output 1.1: Results

The SAS System

The CORR Procedure

4 Variables: USStocks USBonds Pacific European

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
USStocks	48	11.83	16.79	567.72	-36.55	37.20

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
USBonds	48	7.50	9.55	360.09	-11.12	32.81
Pacific	48	12.92	29.19	620.35	-36.17	107.55
European	48	12.20	21.45	585.72	-46.08	79.79

Pearson Correlation Coefficients, N = 48
Prob > |r| under H0: Rho=0

	USStocks	USBonds	Pacific	European
USStocks	1.00000	-0.02015	0.44319	0.75717
		0.8919	0.0016	<.0001
USBonds	-0.02015	1.00000	-0.18788	-0.05029
	0.8919		0.2010	0.7343
Pacific	0.44319	-0.18788	1.00000	0.58032
	0.0016	0.2010		<.0001
European	0.75717	-0.05029	0.58032	1.00000
	<.0001	0.7343	<.0001	

Asset return is modeled as random variable X in the modern portfolio theory. The expected value of this random variable $E(X)$ is called expected return. Expected return is often notated using Greek symbol μ in financial literature convention. The variance of this random variable $V(X)$, or the standard deviation $\sigma_x = \sqrt{V(X)}$, is a measurement of risk in financial literature.

Assume that there are n investable assets with expected return vector $\mu' = (\mu_1, \mu_2, \ldots, \mu_n)$ and variance covariance matrix $C = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \cdot & \cdot & \cdot \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}$, where c_{ij} is the covariance of asset i with asset j,

If portfolio P has weight $w' = (w_1, w_2, \ldots, w_n)$ on assets i=1,2,…,n. Then the portfolio's expected return is

$$\mu_p = \sum_{i=1}^{n} w_i \mu_i \qquad (5)$$

And the portfolio's variance is

$$V_p = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j c_{ij} \qquad (6)$$

1.6 Efficient Portfolio

The decision variables in portfolio theory are portfolio weights. Markowitz (1952, 1959) created Modern Portfolio Theory, often denoted as MPT, and stipulated that portfolio weights should be chosen such that the portfolio is mean-variance efficient. In other words, no other portfolio has higher expected returns with the same risk and no other portfolio has lower risk with the same expected return. Mean-variance efficient portfolios are also called *efficient frontier*. The efficient frontier can be found by quadratic programming

$$\min_w V_p \tag{7a}$$

Such that

$$\mu_p = E_p \tag{7b}$$

$$Aw = b \tag{7c}$$

$$w \geq 0 \tag{7d}$$

where constraint (7b) is portfolio's expected return and (7d) is no-short selling constraint, i.e. every weight w_i must be nonnegative, and constraint (7c) is used to make sure the portfolio has the desired characteristics like industry exposures and factor exposures. A is an $m \times n$ matrix, and b is an m component vector. There is no analytic solution in general. In textbook portfolio theory, the only linear constraint of (7c) is usually budget constraint

$$\sum_{i=1}^{n} w_i = 1 \tag{8}$$

and the short-selling constraint (7d) is often ignored. In this case,

the efficient portfolio can be found by the following unconstrained optimization problem

$$\min_w V_p - \lambda_\mu \left(\sum_{i=1}^{n} w_i \mu_i - E_p\right) - \eta\left(\sum_{i=1}^{n} w_i - 1\right) \tag{9}$$

where Lagrange multiplier λ_μ is called risk return trade-off parameter in financial literature.

By taking the derivative with respect to all weight variable w_i, the first order conditions of (9) in matrix form is

$$2Cw = \lambda_\mu \mu + \eta \ell$$

where ℓ is the vector of ones, i.e. $\ell' = (1,1,\ldots,1)$. This implies the optimal portfolio weight has the general form

$$w = 0.5\lambda_\mu C^{-1}\mu + 0.5\eta C^{-1}\ell \tag{10}$$

Together with expected return constraint (7b) and budget constraint (8), the efficient portfolio is

$$w = \frac{E_p \ell' C^{-1}\ell - \mu' C^{-1}\ell}{\mu' C^{-1}\mu \ell' C^{-1}\ell - (\mu' C^{-1}\ell)^2} C^{-1}\mu + \frac{\mu' C^{-1}\mu - E_p \mu' C^{-1}\ell}{\mu' C^{-1}\mu \ell' C^{-1}\ell - (\mu' C^{-1}\ell)^2} C^{-1}\ell \tag{11}$$

The variance of the corresponding portfolio is;

$$V_p = \left(\frac{E_p \ell' C^{-1}\ell - \mu' C^{-1}\ell}{\mu' C^{-1}\mu \ell' C^{-1}\ell - (\mu' C^{-1}\ell)^2}\right)^2 \mu' C^{-1}\mu + \left(\frac{\mu' C^{-1}\mu - E_p \mu' C^{-1}\ell}{\mu' C^{-1}\mu \ell' C^{-1}\ell - (\mu' C^{-1}\ell)^2}\right)^2 \ell C^{-1}\ell$$

$$+2\frac{E_p \ell' C^{-1}\ell - \mu' C^{-1}\ell}{\mu' C^{-1}\mu \ell' C^{-1}\ell - (\mu' C^{-1}\ell)^2}\frac{\mu' C^{-1}\mu - E_p \mu' C^{-1}\ell}{\mu' C^{-1}\mu \ell' C^{-1}\ell - (\mu' C^{-1}\ell)^2}\mu' C^{-1}\ell \qquad (12)$$

The simple constrained portfolio optimization problem does have an analytical solution.

1.7 Minimum Variance Portfolio

Let us study a special portfolio with the expected return set to

$$E_{\min} = \frac{\mu' C^{-1}\ell}{\ell' C^{-1}\ell} \qquad (13a)$$

The corresponding portfolio weight vector is

$$w_{\min} = \frac{C^{-1}\ell}{\ell' C^{-1}\ell} \qquad (13b)$$

The variance of this portfolio is

$$V_{\min} = \frac{1}{\ell' C^{-1}\ell} \qquad (13c)$$

Portfolio $w_{\min}$ is called the minimum variance portfolio. It achieves the minimum risk among all the portfolio combinations.

If we choose the portfolio's expected return to be

$$E_p = E_m = \frac{\mu' C^{-1}\mu}{\mu' C^{-1}\ell} \qquad (14)$$

Then the corresponding efficient portfolio weight is

$$w_m = \frac{C^{-1}\mu}{\mu' C^{-1}\ell} \qquad (15)$$

And the variance of this portfolio is

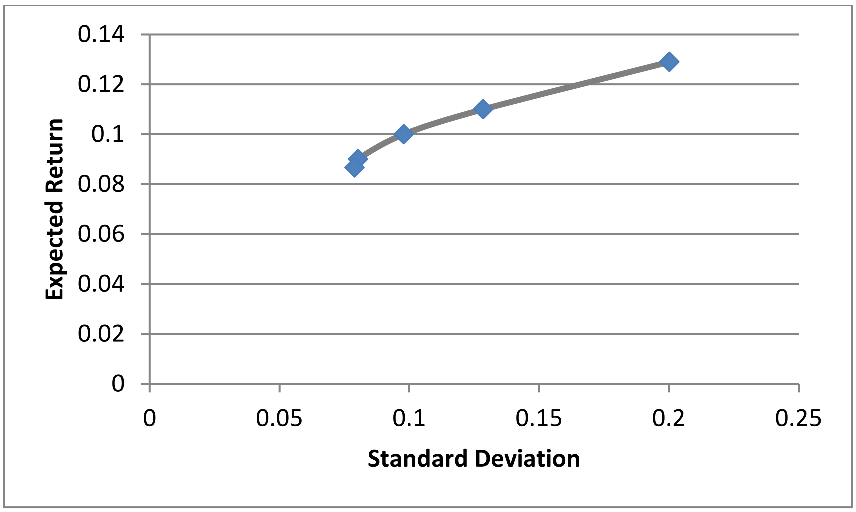
$$V_{\mathrm{m}} = \frac{\mu' C^{-1}\mu}{(\mu' C^{-1}\ell)^2} \qquad (16)$$

1.8 Market Portfolio

The portfolio w_m is called the market portfolio. Equation (11) is a special case of the two-fund theorem, which states that all efficient portfolios are a linear combination of two basic efficient portfolios. Here the two basic efficient portfolios are the minimum variance portfolio and market portfolio. If an investor can borrow and lend at a risk-free rate, then the minimum variance portfolio is the portfolio composed of 100 percent risk-free assets. The two-fund theorem becomes the two-fund separation theory of Tobin (1958).

We can use the sample average return as expected returns vector and the sample covariance as variance covariance matrix to generate global efficient frontier as depicted in Figure 1.3.

Figure 1.3: Global Risk-Return Tradeoff (Efficient Frontier): 1970-2017



The minimum variance portfolio and market portfolio with four asset classes consisting of US stock, US Bond, Pacific Stock, and Europe Stock is presented in Table 1.3.

Table 1.3: Global Minimum Variance and Market Portfolios

Asset Classes	$w_{\min}$	w_m	w_{gm}
US Stock	0.26	0.20	0.79
US Bond	0.66	0.75	0.00
Pacific	0.09	0.08	0.19
Europe	-0.02	-0.03	0.02
Expected Return	0.09	0.09	0.12
Standard Deviation	0.08	0.08	0.16

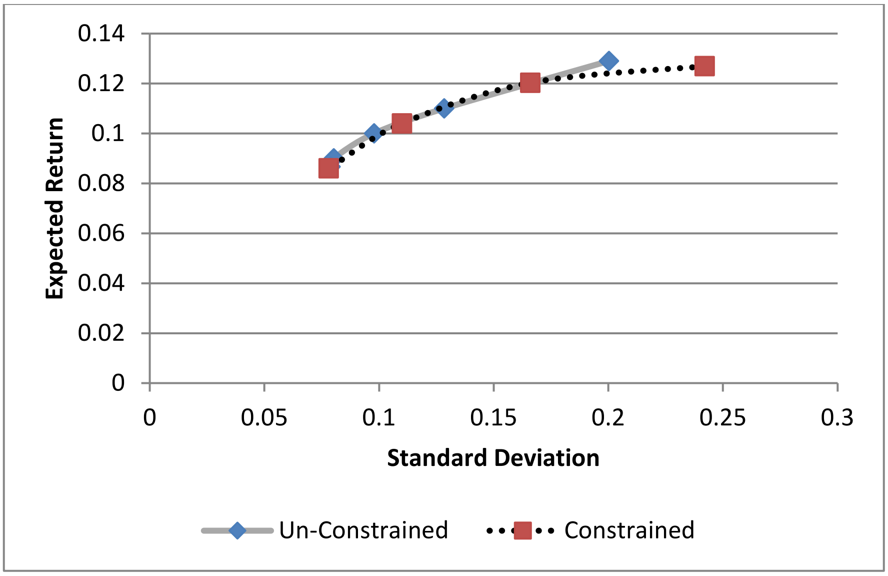
In real-life portfolio management, some form of nonnegative constraint (7d) and affine constraint (7c) is always present.

Markowitz (1959) invented a fast and efficient way – the critical line algorithm – to find all of the efficient portfolios satisfied by the general form of constraint (7c) and (7d) with expected return E_p as a parameter. Block et al (1993) used the critical line algorithm to run hundreds of simulations in the 1990s. The general form of constraint (7c) can handle inequality constraints too by adding auxiliary slack variables.

As expected, the unconstrained efficient frontier dominates the constrained efficient. The difference is small in this four-asset class case. This is foreseen by Table 1.3 in which both the minimum variance portfolio and market portfolio are short selling only two and three percent of Europe stock. Two or three percentage differences in weights does not change the portfolio's mean and risk significantly.

We have used sample means as expected returns and the sample variance-covariance matrix as input to generate efficient portfolios arising from the optimization problem of general form (7). In Table 1.3 we report a geometric mean maximizing portfolio w_{gm} , which is 79 percent US stock, 19 percent Pacific stock, and two percent European stock. This is the second portfolio from the right reported on the constrained Efficient Frontier in Figure 1.4. Portfolio w_{gm} achieves the maximum geometric mean.

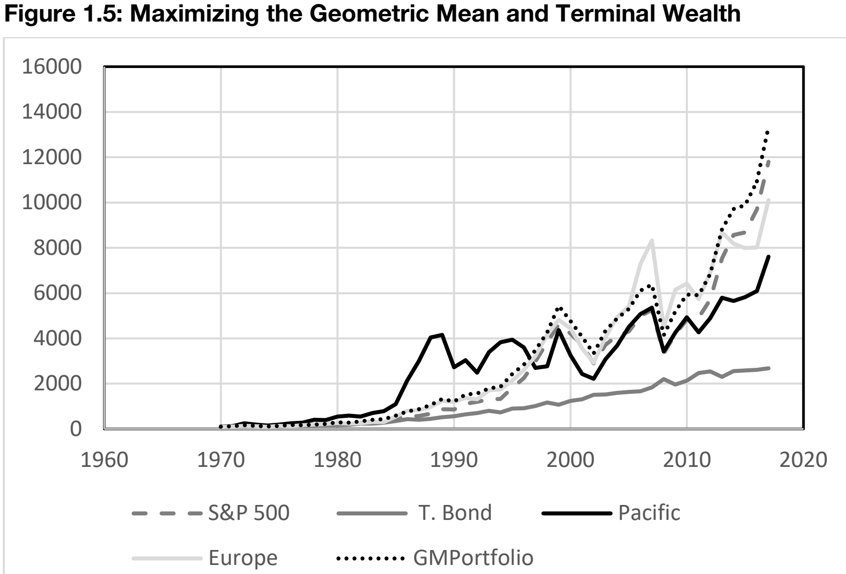
Figure 1.4: Comparison of Constrained vs Unconstrained Risk-tradeoff curves.



1.9 Portfolio Optimization

The mean-variance portfolio analysis enables investors to choose the best portfolio to suit their risk tolerance. Retirees do not have to invest all money in bond. The minimum variance portfolio $w_{\min}$, with two-thirds of the fund in bonds and one-third of the fund in stocks, is as low risk as buying 100 percent US bonds while making one hundred basis points more return. Young people who are ages 35 and younger and who would like to grow their funds should invest in geometric mean maximizing portfolio w_{gm}. Figure 1.5 shows the cumulative effect of optimal portfolio combinations.

Figure 1.5: Maximizing the Geometric Mean and Terminal Wealth

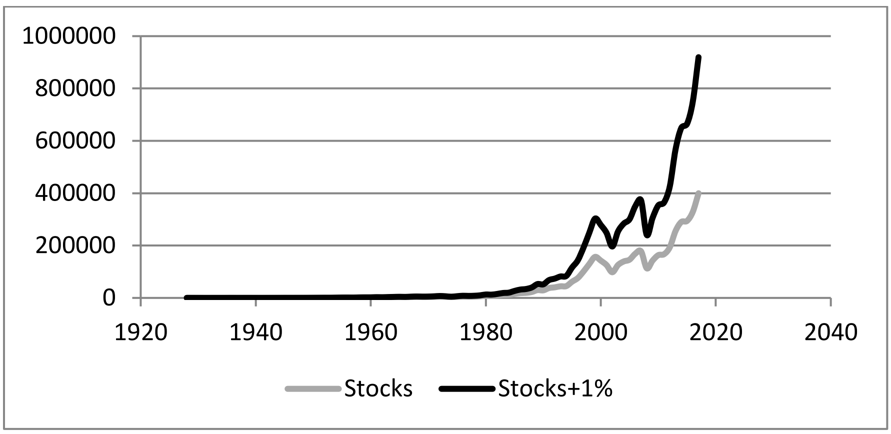


As shown in Figure 1.5, $100 starting in 1970 turns into $13,350.68 for investing in the gm-maximizing portfolio while US stock turns into $11,799.48. This is a great enhancement in terminal wealth!

This book explains how to build expected returns of a thousand securities as portfolio optimization input. The past average return contains only part of the information related to future expected return. There is also future-related information in the fundamental variables, see Block et al (1993) and later chapters of this book. We will introduce financial statements and financial ratios in Chapter 2, and then teach how to build expected return models in Chapter 4. Modern Markowitz-based portfolio construction models are explored in Chapter 5.

The goal of active management is to beat market portfolio by a couple of hundred basis points on an annual basis. Figure 1.6 shows the astounding results of beating the market by one hundred basis points. This book shows that it is possible by deploying advanced statistical tools and disciplined scientific portfolio optimization.

Figure 1.6: The Cumulative Effect of Beating the Market



1.10 Summary and Conclusions

In this chapter we introduced the reader to the concept that current savings can be invested into stocks and bonds that can enhance terminal wealth. This is a risk-return concept that is extremely important for investors. Stocks have produced more returns than bonds over the past time periods from 1928 – 2017 and 1970-2017 because their risk, as measured by the standard deviation is greater. The risk-return trade-off does change over time, but if an investor has a 30-year investment horizon, then stocks will be preferred to bonds to maximize terminal wealth. If an investor can earn more than one percent above the market return, then terminal wealth is greatly increased. The purpose of this book is to show readers how use SAS to enhance their wealth.