SECTION 4

The Future

CHAPTER 18. POPULARITY AND PREMIUMS

Thomas M. Idzorek, CFA
Chief Investment Officer, Morningstar

Roger G. Ibbotson, PhD
Professor in the Practice Emeritus of Finance, Yale School of Management
Chair, Zebra Capital Management, LLC

Paul D. Kaplan, PhD, CFA
Director of Research (Retired), Morningstar Canada

Introduction

The analysis of equity returns has, in large part, concentrated on identifying common factors in individual security returns. Factors already discussed in Chapters 2, 4, and elsewhere include small size, value, and illiquidity, as well as the "market factor," or beta factor, which all stocks have in common to a greater or lesser degree.

This chapter looks at the root cause of factors, identified and not yet identified, through the lens of a formal demand-side asset pricing theory that we call the popularity asset pricing model (PAPM). It addresses some of the weaknesses of the capital asset pricing model (CAPM) and considers all the asset characteristics that might make an investor like or dislike a security, affecting its price and thus its expected return. Summed up in one sentence, the PAPM says that assets with popular characteristics have lower expected returns because their popularity has caused them to have higher prices.

Popularity is not a "factor" in and of itself; rather, it is what drives demand—the collective, aggregated, wealth-weighted preferences of investors. It is the attribute that links together all possible factors—those that a given investor prefers or likes and those to which the investor is averse or that the investor dislikes. Aggregating across investors, the PAPM reflects the demand for a security and thus falls into the category of demand-side models of expected return, as described later.

The PAPM itself says nothing *a priori* about whether investors *should* like or dislike a characteristic; it only asks whether they *do* like or dislike it. Other information, however, can be used to determine whether a characteristic is likely to be regarded by investors as "good" or 'bad."

Some asset characteristics are almost universally liked or disliked. For instance, investors like expected returns and dislike risk; expected returns are popular, and risk is not. These two deep-seated investor *preferences*—desiring expected return and disliking risk—are the fundamental drivers of the risk–expected return paradigm. The tension between them explains much about the risk premiums discussed in this book.

But there is more to the story. Investor preferences—their likes and dislikes—go beyond risk and expected return, are nuanced, and evolve over time. It is the collective actions of investors acting on their return expectations and preferences that determine asset prices and create *priced* factors with expected return premiums, as well as other factors that help explain realized stock returns but that do not offer expected return premiums.

Chapter 2 highlights the various *realized* "risk" premiums measured in 1976 by Roger Ibbotson and Rex Sinquefield and in later work by Ibbotson with various coauthors—an equity premium, a small-cap premium, a value premium, a liquidity premium, a horizon premium, and a default premium. The assumption that the premiums persist was the basis for forecasting the future via simulations. But why should such premiums persist?

Research by Roger Ibbotson, Thomas Idzorek, Paul Kaplan, and James Xiong provides an answer: the theory of popularity and, as we noted at the outset, its formal model, the popularity asset pricing model.

In this chapter, we review the theory of popularity and the creation of the PAPM. Looking backward, we examine various realized premiums and discounts through the popularity lens. Looking forward, we apply the theory of popularity and provide our forecasts or express confidence about which premiums are likely to persist into the future.

The Theory of Popularity

The theory of popularity provides a plausible explanation for most realized historical premiums and, as such, is a powerful framework for thinking about which premiums can be expected to persist over the long term. A wide variety of risk-related and non-risk-related "characteristics" is associated with securities, investments, or companies that investors *like* or *dislike*, all of which can influence investor portfolio formation and thus asset prices.

If enough investors *like* a particular characteristic, investments with that characteristic will tend to be in *high* demand or *popular*, which makes them relatively *expensive*, providing new purchasers at that *high* price with *lower* expected returns. Conversely, if enough investors *dislike* a particular characteristic, investments with that characteristic will tend to be in *low* demand or *unpopular*, which makes them relatively *inexpensive*, providing new purchasers at that *low* price with *higher* expected returns. Of course, any given investment is a bundle of a wide variety of characteristics—some that investors like and others that they dislike.

In the 1980s, in a Graham and Dodd Scroll Award–winning article in the *Financial Analysts Journal*, "The Demand for Capital Market Returns: A New Equilibrium Theory," the seeds of what would become the theory of popularity and the PAPM were first sown by Roger Ibbotson, Jeffrey Diermeier, and Laurence Siegel (1984). They argued that investors care about characteristics beyond risk, which include liquidity and taxation.

In the 2010s, Idzorek, Xiong, and Ibbotson (2012) and, in another Graham and Dodd Scroll Award–winning article, Ibbotson, Chen, Kim, and Hu (2013) documented liquidity as a key characteristic that investors like. Looking at either funds or individual securities, portfolios with lower liquidity outperformed portfolios with higher liquidity. High liquidity is unambiguously popular, leading to a liquidity discount or illiquidity premium. The CAPM, the dominant asset pricing model, is silent on liquidity.

Seeking a better asset pricing theory, Ibbotson and Idzorek (2014) and Idzorek and Ibbotson (2017) built on and provided additional rationales for the New Equilibrium Theory, presenting the theory of popularity as a broad, unifying investor-preference-driven explanation for observed anomalies and premiums. In a CFA Institute Research Foundation book, *Popularity: A Bridge Between Classical and Behavioral Finance*, Ibbotson, Idzorek, Kaplan, and Xiong (2018) built on the previous work, put forth new empirical evidence supporting popularity, and, thanks largely to Paul Kaplan, put forth a formal equilibrium model: the PAPM.

Most recently, in a Harry M. Markowitz Award–winning article in the *Journal of Investment Management*, Idzorek, Kaplan, and Ibbotson (2025) expanded the PAPM to allow investors to *disagree* on the future payoffs of assets. In doing so, the PAPM addresses what Fama and French (2007) identified as the two most significant weaknesses of the CAPM. In their article "Disagreement, Tastes, and Asset Prices," Fama and French identified as the two key weaknesses of the CAPM (1) the assumption that all investors "agree" on the distributions of future payoffs of all assets (i.e., no disagreement) and (2) the assumption that the only investor "taste" (what we call preference) is risk tolerance.

The PAPM addresses both weaknesses, simultaneously allowing for disagreement and any number of investor tastes or preferences. In contrast with the CAPM, in which each investor solves the Markowitz mean–variance problem (although that is actually unnecessary given that all investors "agree" on the capital market assumptions), in the PAPM, each investor solves an expanded Markowitz mean–variance problem that includes an additional term capturing their nonpecuniary preferences.

In both equilibrium models—the CAPM and the PAPM—it is the collective, aggregated, wealth-weighted actions of investors that lead to equilibrium prices. With the CAPM, all investors agree on expected returns and covariances, and the only preference is risk tolerance. With the PAPM, each investor creates a *personalized* portfolio based on the agreed-on covariance matrix of returns, their personal expected returns, and any number of additional preferences for characteristics.[1] The CAPM is actually a special and highly unrealistic case of the PAPM, in which investors all agree on expected returns and the only preference is risk tolerance.

Given that markets are relatively informationally efficient, disagreements that lead to outperformance and underperformance are generally short lived and do not result in expected premiums. In contrast, a number of investor preferences are relatively permanent (e.g., disliking risk, preferring more liquidity, and disliking taxes) and thus, from our perspective, are likely to result in expected long-term premiums.

Of course, changes in investor preferences can occur that lead to short-term movement around long-term mean expectations. One example is the desire for liquidity during a market crisis: As a crisis begins, investors pay up for liquidity, and as the crisis ends, the gains of the run-up are ceded. Another example of changes in investor preferences might be the recent period when environmental, social, and governance (ESG) investing increased in relative popularity with regulatory tailwinds and then became less popular as it became highly politicized and regulatory tailwinds turned into headwinds. Looking forward, regulatory changes can cause large-scale shifts in investor demand. For example, allowing retail investors to purchase private equity and credit funds could shift the demand (popularity) for private funds.

---

[1] For more information on personalized portfolio construction based on the PAPM, see Idzorek and Kaplan (2022, 2024) and Idzorek (2023).

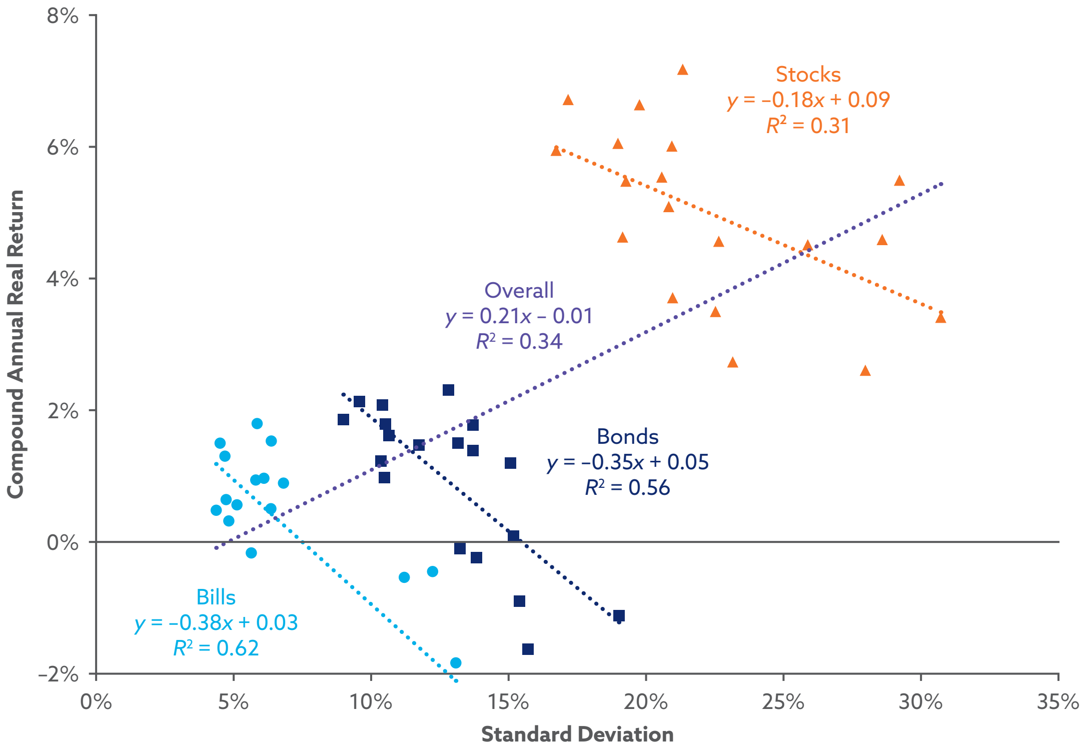
Our primary interest is to understand how popularity provides an explanation for various realized premiums. We then use the theory of popularity to look ahead and identify which premiums are most likely to persist into the future.

Realized Premiums from a Popularity Perspective

Investors dislike risk. This is arguably the most powerful investor preference and, in our opinion, the most plausible explanation for why the risk–return paradigm tends to hold *across* asset classes with different inherent levels of risk. This phenomenon was abundantly clear from the 100-year wealth charts in Chapter 2, in which US bonds beat US bills, US large-cap stocks beat US bonds, and US small-cap stocks beat US large-cap stocks. *Within* an asset class in which investments tend to have similar levels of risk, additional investor preferences help explain realized premiums.

This principle is perhaps best illustrated by leveraging the 126-year history documented by Elroy Dimson, Paul Marsh, and Mike Staunton (2002) in *Triumph of the Optimists*, updated through 2025 in UBS's *Global Investment Returns Yearbook 2026* (Dimson, Marsh, and Staunton 2026) and summarized in Chapter 11 of this book. In **Exhibit 18.1**, we plot the realized compound

Exhibit 18.1. Risk and Realized Real Return of Stocks, Bonds, and Bills for 19 Countries, 1900–2025



Sources: Updated analysis from Ibbotson et al. (2018). Author calculations based on data from Dimson et al. (2026).

Notes: The 19 countries are Australia, Belgium, Canada, Denmark, Germany, Ireland, Japan, the Netherlands, Finland, France, Italy, Norway, New Zealand, South Africa, Spain, Sweden, Switzerland, the United Kingdom, and the United States. There were no returns on bonds or bills for Germany for 1922 and 1923, so those years for Germany are omitted.

annual real returns (in local currencies) and standard deviations associated with stocks, bonds, and bills for 19 countries.[2] Based on all 57 plot points, the best-fit regression line has a positive slope of 0.21 and an R^2 of 0.34. In other words, there is a moderately positive relationship between risk and realized return across the 57 asset classes, but it is far from a perfect fit.

Notice that we color-coded the plot points for stocks (orange triangles), bonds (dark blue squares), and bills (light blue dots), leading to three distinct asset class color clusters. We then calculated three more best-fit regression lines for the three asset classes, each of which shows a negative risk–return relationship!

So, although risk is the dominant characteristic that investors dislike *across* asset classes, *within* asset classes that tend to have similar risk characteristics it is the nuanced collective preferences of investors that lead to anomalies relative to what standard asset pricing models would have predicted (see Idzorek et al. 2025).

Turning to some of the most well-known anomalies and *historical* premiums and discounts, in **Exhibit 18.2** we list various realized premiums and discounts and their popularity-based explanations.

Exhibit 18.2. Popularity-Based Explanations for Realized Premiums and Discounts

Premium or Discount	Popularity-Based Explanation
Equity premium	Investors dislike: risk (prefer safety)
Illiquidity premium	Investors dislike: illiquidity
Small-cap premium	Investors dislike: risk (prefer safety), illiquidity, information costs
Value premium	Investors dislike: boring, nonglamorous; like the opposite
Low-beta premium	Active fund managers prefer high beta to outperform benchmarks
Severe downside risk premium	Investors dislike: severe risk (prefer safety)
ESG/green discount	Investors like: ESG and green
Good reputation discount	Investors like: good reputations
Credit/default premium	Investors dislike: risk (prefer safety)
Horizon/duration premium	Investors dislike: interest rate risk, reinvestment risk, duration mismatches

[2] The data are for 20 countries, but we omitted Austria because of the extreme volatility of its bond index. Also, there are no returns on bonds and bills for Germany for 1922 and 1923, so for German bonds and bills, we used the remaining 124 years. For all stock indices and for all other bond and bill indices, we used the full 126-year history.

One anomaly omitted from Exhibit 18.2 is momentum. From a popularity perspective, momentum might be explained as a series of short-term changes in relative popularity, in which some event triggers a momentum-building cycle for a particular investment. During the first part of the cycle, investor attention, trading, and liquidity—characteristics that investors tend to like—all increase. At some point, the price rises to a point at which sellers outnumber buyers, and then the price falls. Thus, momentum cannot be a permanently priced factor, but it can have a powerful effect on returns while it is taking place. Momentum is an example of mispricing.

In the next section, we elaborate on the popularity-based explanations of the premiums and discounts, with an eye toward the future.

Forecasting Long-Term Expected Premiums

Researchers have identified hundreds of possible factors—the so-called "factor zoo."[3] In "Domesticating the Factor Zoo with Economic Theory," from the 50th anniversary issue of the *Journal of Portfolio Management*, Idzorek, Kaplan, and Ibbotson (2024) used the PAPM to analyze and identify long-term *priced* factors.

Empirical attempts to tame the factor zoo are helpful, especially from a backward-looking perspective, but a forward-looking perspective should be based on theory. When attempting to domesticate the factor zoo with theory, it is important to distinguish between *priced* factors that lead to expected premiums/discounts and *attribution* (explanatory) factors that do not.

The ever-growing factor zoo includes hundreds of possible attribution factors that at some point helped explain realized return variation among stocks. Attribution factors are numerous, do not emanate from an asset pricing model, and have a zero expected long-term mean; thus, they *should not* influence long-term investment policy.

In contrast, priced factors are quite rare, should emanate from an asset pricing model, and have a nonzero expected long-term mean; thus, they *should* influence long-term investment policy. Through a forward-looking popularity lens, we make conjectures on the premiums and discounts, including whether they will persist into the future.

- *Equity risk premium—high confidence:* Investors uniformly dislike risk, and based on their risk tolerance, a large subset of investors therefore choose to diversify away from equities, avoid equities altogether, or move in and out based on their view of the equity risk premium. To be incentivized to hold equities, based on their respective risk preferences, investors require the expectation of a return premium as compensation for taking on additional risk. We note that even though we have high confidence in the equity risk premium, there is no guarantee that equities will outperform bills or bonds. It is unlikely but possible that equities could underperform for multiple decades. Similar disclaimers apply to all the *expected* premiums because no premium is guaranteed.
- *Illiquidity premium—high confidence:* Investors uniformly dislike illiquidity and like liquidity. A large number of investors will thus continue to avoid less liquid investments, preferring to own more liquid ones. Both of these preferences and corresponding behaviors should continue to create an illiquidity premium.

[3] John Cochrane (2011) used the term "factor zoo" during a 2011 presidential address to the American Finance Association.

- *Small-cap premium—moderate confidence:* To the degree that small-cap stocks are more risky than large-cap stocks, investors require the expectation of a premium. Small caps are less liquid and less well known, with less information than large-cap stocks, which are characteristics that investors dislike and contribute to a "bundled" small-cap premium.

  At least three practices counteract a small-cap premium. First, the widespread belief in a small-cap premium leads to a group of "small-cap premium harvesters," whose collective action bids up the prices of these stocks and thus mutes the premium. Second, private companies that choose to stay private decrease the supply of public small-cap companies, which (all else equal) makes small-cap companies relatively expensive. Third, many quant managers overinvest in small-cap stocks because they often equally weight their portfolios.
- *Value premium—low confidence:* Between the discovery of the value premium by Basu (1977) and the subsequent publication of Fama and French (1992, 1993), which made the existence of the premium in historical data more widely known, the value premium largely disappeared. We attribute the value premium to both disagreement and tastes/preferences, with a group of investors systematically overestimating the growth and profitability expectations for glamorous growth stocks and another large group of investors simply preferring glamorous, high-growth investments. The discovery of the value premium gave rise to "value premium harvesters," whose collective actions dampen the magnitude of this premium. We believe investors will continue to overestimate the future growth and profitability of growth stocks, with a partially overlapping group of investors continuing to prefer growth stocks for other reasons. As has been the case since the early 1990s, value premium harvesters will mute the value premium moving forward.
- *Severe downside premium—high confidence:* Investors uniformly dislike risk and especially dislike severe downside risk. A large number of investors will thus continue to avoid investments perceived to have severe downside risk. This avoidance should continue to create a severe downside risk premium.
- *Low-beta premium—low confidence:* From a popularity-based perspective, the low-beta anomaly, in which low-beta stocks tend to outperform high-beta stocks on a risk-adjusted basis, results from active managers seeking to outperform their benchmarks without using leverage. Such managers overweight high-beta stocks as an imperfect substitute for holding a leveraged portfolio. Xiong, Idzorek, and Ibbotson (2023) documented the dramatic and steady decrease in active mutual funds as a percentage of the market. We expect this trend to continue, suggesting that the low-beta anomaly should decrease and might disappear. Leverage aversion will continue to contribute to a small premium.
- *ESG/green discount—moderate confidence:* ESG/green investing and sustainability have unfortunately been politicized. All else equal, a large group of investors prefer investments that appear to be ESG-friendly or "greener." New investors will thus pay a premium price—that is, they will receive lower expected returns. Conversely, investors willing to purchase investments with bad ESG characteristics should realize a premium.
- *Brand/reputation discount—moderate confidence:* Similar to the ESG/green discount, all else equal, a large group of investors prefers investments with strong brands and/or reputations. Investors will therefore pay a premium price for such stocks and will thus receive lower expected return. We call this effect on return the brand/reputation discount. Conversely, those willing to purchase investments with poor brands and reputations should realize a premium.

- *Credit/default premium—high confidence:* Analogous to the equity premium, investors uniformly dislike risk, causing a large subset of investors to diversify away from credit risk. To be motivated to hold bonds with higher levels of credit risk, investors require a yield in excess of their expected default losses to compensate them for taking on default risk.
- *Horizon/duration premium—mixed confidence, depending on horizon:* We are confident that bonds will continue to provide a horizon/duration premium over bills. Within the bond market, however—because of large demand from institutional investors seeking to match the duration of their liabilities, which are often long term—we have less confidence that long-duration bonds will provide a significant premium over intermediate-duration bonds.

Key Takeaways

- The PAPM is a demand-side model in which the aggregated, wealth-weighted, personalized portfolio construction decisions of investors, based on their expected returns and any number of risk- and non-risk-related preferences, collectively lead to equilibrium asset prices.
- The theory of popularity helps us understand the realized premiums of the past, differentiate between *attribution* factors and *priced* factors, and predict which priced factors are likely to persist.
- Disliked characteristics, such as high risk, low liquidity, long duration, low growth, boring businesses, and polluters, are unpopular, causing stocks with these characteristics to have relatively low current prices and relatively higher expected returns.
- Conversely, liked characteristics, such as low risk, high liquidity, short duration, high growth, glamorous businesses, and a green environmental profile, are popular, causing stocks with these characteristics to have relatively high current prices and relatively lower expected returns.

References

Basu, Sanjoy. 1977. "Investment Performance of Common Stocks in Relation to Their Price–Earnings Ratios: A Test of the Efficient Market Hypothesis." *Journal of Finance* 32 (3): 663–82. doi:10.1111/j.1540-6261.1977.tb01979.x.

Cochrane, John H. 2011. "Presidential Address: Discount Rates." *Journal of Finance* 66 (4): 1047–108. doi:10.1111/j.1540-6261.2011.01671.x.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2002. *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton, NJ: Princeton University Press.

———. 2026. *Global Investment Returns Yearbook 2026*. UBS Global Wealth Management. www.ubs.com/global/en/wealthmanagement/insights/2026/global-investment-returns-yearbook.html.

Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47 (2): 427–65. doi:10.1111/j.1540-6261.1992.tb04398.x.

———. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56. doi:10.1016/0304-405X(93)90023-5.

———. 2007. “Disagreement, Tastes, and Asset Prices.” *Journal of Financial Economics* 83 (3): 667–89. doi:10.1016/j.jfineco.2006.01.003.

Ibbotson, Roger G., Jeffrey J. Diermeier, and Laurence B. Siegel. 1984. “The Demand for Capital Market Returns: A New Equilibrium Theory.” *Financial Analysts Journal* 40 (1): 22–33. doi:10.2469/faj.v40.n1.22.

Ibbotson, Roger G., and Thomas M. Idzorek. 2014. “Dimensions of Popularity.” Special 40th anniversary issue, *Journal of Portfolio Management* 40 (5): 68–74. doi:10.3905/jpm.2014.40.5.068.

Ibbotson, Roger G., Thomas M. Idzorek, Paul D. Kaplan, and James X. Xiong. 2018. *Popularity: A Bridge Between Classical and Behavioral Research*. CFA Institute Research Foundation (10 December). https://rpc.cfainstitute.org/research/foundation/2018/popularity-bridge-between-classical-and-behavioral-finance.

Ibbotson, Roger G., and Rex A. Sinquefield. 1976. “Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926–1974).” *Journal of Business* 49 (1): 11–47. doi:10.1086/295803.

Ibbotson, Roger G., Zhiwu Chen, Daniel Y.-J. Kim, and Wendy Y. Hu. 2013. “Liquidity as an Investment Style.” *Financial Analysts Journal* 69 (3): 30–44. https://doi.org/10.2469/faj.v69.n3.4.

Idzorek, Thomas M. 2023. “Personalized Multiple Account Portfolio Optimization.” *Financial Analysts Journal* 79 (3): 155–70. doi:10.1080/0015198X.2023.2212581.

Idzorek, Thomas M., and Roger G. Ibbotson. 2017. “Popularity and Asset Pricing.” *Journal of Investing* 26 (1): 46–56. doi:10.3905/joi.2017.26.1.046.

Idzorek, Thomas M., and Paul D. Kaplan. 2022. “Forming ESG-Oriented Portfolios: A Popularity Approach.” Special 30th anniversary issue, *Journal of Investing* 31 (4): 63–75. doi:10.3905/joi.2022.31.4.063.

———. 2024. *Lifetime Financial Advice: A Personalized Optimal Multilevel Approach*. Charlottesville, VA: CFA Institute Research Foundation. https://rpc.cfainstitute.org/research/foundation/2024/lifetime-financial-advice-a-personalized-optimal-multilevel-approach.

Idzorek, Thomas M., Paul D. Kaplan, and Roger G. Ibbotson. 2024. “Domesticating the Factor Zoo with Economic Theory.” Special 50th anniversary issue, *Journal of Portfolio Management* 51 (1): 97–107. doi: 10.3905/jpm.2024.51.1.097.

———. 2025. “The CAPM, APT, and PAPM.” *Journal of Investment Management* 23 (1): 39–55. https://joim.com/article/the-capm-apt-and-papm/.

Idzorek, Thomas M., James X. Xiong, and Roger G. Ibbotson. 2012. “The Liquidity Style of Mutual Funds.” *Financial Analysts Journal* 68 (6): 38–53. doi:10.2469/faj.v68.n6.3.

Xiong, James X., Thomas M. Idzorek, and Roger G. Ibbotson. 2023. “The Diminishing Role of Active Mutual Funds: Flows and Returns.” *Journal of Investment Management* 21 (4): 32–52. https://joim.com/article/the-diminishing-role-of-active-mutual-funds-flows-and-returns.

CHAPTER 19. PARAMETRIC FORECASTING

Paul D. Kaplan, PhD, CFA
Director of Research (Retired), Morningstar Canada

In 1976, Ibbotson and Sinquefield published a pair of articles in the *Journal of Business* that would have great impact on our understanding of capital markets. In the first (Ibbotson and Sinquefield 1976a), they documented monthly returns on stocks, bonds, and bills, as well as monthly inflation rates, over the period 1926–1974. In the second (Ibbotson and Sinquefield 1976b), they applied a bootstrap method to the historical data to make probabilistic forecasts about cumulative wealth over several decades into the future. In Chapter 20, Roger G. Ibbotson, William N. Goetzmann, and Otto S. Manninen apply this bootstrap method using 100 years of data.

Lewis, Kassouf, Brehm, and Johnston (1980) demonstrated how the results of Ibbotson and Sinquefield (1976b) can largely be replicated using a parametric model based on the lognormal distribution. Their lognormal forecasting model has only two inputs: the mean and standard deviation of return. These are the first two moments of any distribution. Using only mean and standard deviation raises the question, Would it make a difference if higher moments—namely, the third (skewness) and the fourth (excess kurtosis)—are taken into account? The answer to this question depends on how far off the higher moments are from those of a lognormal distribution at long horizons. For well-diversified portfolios, the lognormal model works quite well and is the model of choice. As I demonstrate in this chapter, however, there are cases in which the lognormal model largely understates long-term left-tail risk. In such cases, we need another forecasting model.

In this chapter, I present the lognormal forecasting model and illustrate it using large-cap stocks. I then introduce a new forecasting model that takes into account all four moments and compares the results of the two models using large-cap, midcap, and small-cap stocks. Because this model is based on Johnson (1949) distributions, I call it the Johnson forecasting model.[1] I will show that although the lognormal and Johnson models are nearly indistinguishable for large-cap, midcap, and small-cap stocks, they clearly differ for low percentiles at long horizons. For small-cap stocks, these differences are especially large. In this way, the Johnson model captures the long-term left-tail risk of midcap and small-cap stocks not captured by the lognormal model.

Another kind of parametric forecasting is the formation of capital market assumptions for mean–variance optimization. Assuming an annual rebalancing period, these consist of the expected annual return on each asset class, the annualized standard deviation of return for each asset class, and the correlation of annualized returns between each pair of asset classes. Although the standard deviations and correlations are usually estimated using historical return data, it is important that the expected returns be calibrated to an internally consistent model. I will show how do this at the end of this chapter.

---

[1] Johnson distributions are different from stable Paretian distributions. The moments of a Johnson distribution are finite, whereas the second and higher moments of a stable Paretian distribution are infinite. See Kaplan (2012a, 2012b) to see how stable Paretian distributions can be used to model asset class returns.

In this chapter, I will cover the following topics:

- Estimating the first four monthly moments
- Probabilistic wealth forecasting
- The lognormal forecasting model
- The first four moments of wealth distributions
- The Johnson forecasting model
- Annualizing the first four moments, covariances, *and* correlations
- Calibrating expected returns for mean–variance optimization

Estimating the First Four Monthly Moments

I refer to mean, standard deviation, skewness, and excess kurtosis collectively as the first four moments. **Exhibit 19.1** presents the mathematical definitions of the first four moments and the

Exhibit 19.1. Formulas for Estimating the First Four Moments of Monthly Returns

Moment	Definition	Estimator
Mean (ER stands for expected return)	$ER_1 = \mathrm{E}[\tilde{R}_1]$	$\widehat{ER}_1 = \frac{1}{N}\sum_{i=1}^{N} R_{1i}$
Standard deviation (SD)	$SD_1 = \sqrt{\mathrm{E}[(\tilde{R}_1 - ER_1)^2]}$	$\widehat{SD}_1 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(R_{1i} - \widehat{ER}_1)^2}$
Skewness	$s_1 = \frac{\mathrm{E}[(\tilde{R}_1 - ER_1)^3]}{SD_1^3}$	$\hat{s}_1 = \frac{N}{(N-1)(N-2)}\sum_{i=1}^{N}\left(\frac{R_{1i} - \widehat{ER}_1}{\widehat{SD}_1}\right)^3$
Excess kurtosis	$\kappa_1 = \frac{\mathrm{E}[(\tilde{R}_1 - ER_1)^4]}{SD_1^4} - 3$	$\hat{\kappa}_1 = \frac{N(N+1)}{(N-1)(N-2)(N-3)}\sum_{i=1}^{N}\left(\frac{R_{1i} - \widehat{ER}_1}{\widehat{SD}_1}\right)^4 - \frac{3(N-1)^2}{(N-2)(N-3)}$
$\tilde{R}_1$ = a random monthly return R_{1i} = historical return in month i N = number of months in sample		

Note: Throughout this chapter, numerical subscripts indicate the number of months in the period of a variable or a parameter. In this exhibit, since all the parameters and variables are monthly, the subscript for the number of months in the period is 1.

formulas for estimating them from historical monthly returns.[2,3] **Exhibit 19.2** presents the estimated values of the first four moments of nine asset class indices presented in this book, using all 1,200 months of real (inflation-adjusted) return data on those indices.

An important caveat in using the formulas in Exhibit 19.1 is that they assume that over the entire 1,200-month sample period, the real returns on each asset class are serially independent and drawn from the same distribution. This assumption could be violated in at least three ways:

- The returns could be serially correlated. The last column of Exhibit 19.2 shows the monthly serial correlation coefficient for each of the nine asset class indices. From these results, we see that serial correlation is a major concern only for the cash indices.
- Conditional variance could be dynamic (that is, the variance itself varies over time), as in the GARCH model of Bollerslev (1986).
- The unconditional distribution changes over time. In other words, the monthly distribution is nonstationary.

Exhibit 19.2. Estimates of the First Four Moments of Monthly Real Asset Class Indices, January 1926–December 2025

Asset Class	Mean	Standard Deviation	Skewness	Excess Kurtosis	Serial Correlation
Large-cap stocks	0.69%	5.10%	0.17	6.94	0.08
Midcap stocks	0.83%	6.35%	0.86	12.01	0.13
Small-cap stocks	0.89%	7.41%	1.36	14.77	0.16
US 20-year Treasury bonds	0.19%	2.73%	0.40	3.31	0.11
US 10-year Treasury bonds	0.18%	2.03%	0.37	3.26	0.13
US 5-year Treasury bonds	0.16%	1.43%	0.20	4.06	0.20
US 1-year Treasury bills	0.09%	0.66%	–0.35	7.56	0.43
US 6-month Treasury bills	0.07%	0.57%	–0.93	9.81	0.46
US 30-day Treasury bills	0.03%	0.53%	–1.35	13.54	0.50

Source: Calculations by the author using Ibbotson Indices data.

[2] Ideally, the estimator for mean return should be based on current market conditions, not solely on historical returns. Ibbotson, Goetzmann, and Manninen do this in Chapter 20. In this chapter, to illustrate parametric forecasting, I use historical averages.

[3] Throughout this chapter, I follow the standard practice in probability theory and statistics of using a tilde (~) as part of the name of a random variable and using a circumflex (^) as part of the name of an estimator.

It is possible that one or more of these problems applies to the return histories that form the inputs to the forecasting models presented in this chapter. In particular, dynamic conditional variance and nonstationarity could be the source of high levels of measured excess kurtosis. Using inflation-adjusted returns should somewhat mitigate these issues.

Because the bootstrap model presented in Chapter 20 has results similar to those of the parametric models presented here, however, these issues may not be a major concern. So, for the purposes of the forecasting models, I assume that the monthly returns of each asset class are serially independent and identically distributed and have the first four moments shown in Exhibit 19.2.

Probabilistic Wealth Forecasting

Suppose that at time 0, an investor has W_0 in wealth. Each month, she holds the same portfolio, which she rebalances and holds without additions or withdrawals. After T months, her wealth will be

$$\tilde{W}_T = W_0 \prod_{t=1}^{T} (1+\tilde{R}_{1t}), \tag{19.1}$$

where $\tilde{R}_{1t}$ is the return in month t.

A probabilistic wealth forecast is a percentile of the distribution of $\tilde{W}_T$.[4] Let p be a probability—so, a number between 0% and 100%. $W_T(p)$ is said to be the 100pth percentile of $\tilde{W}_T$ if[5]

$$\Pr[\tilde{W}_T \le W_T(p)] = p. \tag{19.2}$$

Compound annual return (CAR) for T months is

$$\widetilde{CAR}_T = \left(\frac{\tilde{W}_T}{W_0}\right)^{\frac{12}{T}} - 1. \tag{19.3}$$

Therefore, percentiles of CAR can be calculated from wealth percentiles:

$$\widetilde{CAR}_T(p) = \left(\frac{\tilde{W}_T[p]}{W_0}\right)^{\frac{12}{T}} - 1. \tag{19.4}$$

The Lognormal Forecasting Model

Based on the central limit theorem, Lewis et al. (1980) showed how to form probabilistic forecasts using the lognormal distribution. (See the "Central Limit Theorem and the Lognormal Distribution" sidebar.) A monthly return distribution is said to be lognormal if $\log(1+\tilde{R}_1)$ follows a normal distribution. Let μ_1 and σ_1 be the expected value and standard deviation of this

[4] A percentile is a value (1 = nearly the lowest, 99 = nearly the highest) for which a random variable falls below with a specified probability. For example, in Exhibit 19.4, the 75th percentile of wealth per dollar invested at 20 years is 6.50. This means that at 20 years, there is a 75% probability that wealth per dollar invested is 6.50 or less.

[5] In decimal form, p is between 0 and 1. So, the percentile number is 100p. For example, if p = 5% = 0.05, the percentile number is 100p, which equals 5 for the 5th percentile.

normal distribution. These parameters can be calculated from the mean and standard deviation of $\tilde{R}_1$ as follows:

$$\sigma_1 = \sqrt{\log\left[1+\left(\frac{SD_1}{1+ER_1}\right)^2\right]}. \tag{19.5a}$$

$$\mu_1 = \log(1+ER_1) - \frac{1}{2}\sigma_1^2. \tag{19.5b}$$

The 100*p*th percentile of wealth in the lognormal model is

$$W_T(p) = W_0 \exp[T\mu_1 + \sqrt{T}\sigma_1 z(p)], \tag{19.6}$$

where $z(p)$ is the 100*p*th percentile of a standard normal random variable. **Exhibit 19.3** shows values of $z(p)$ for common percentiles.

The 100*p*th percentile of compound annual return in the lognormal model is

$$CAR_T(p) = \exp\left[12\mu_1 + \frac{12}{\sqrt{T}}\sigma_1 z(p)\right] - 1. \tag{19.7}$$

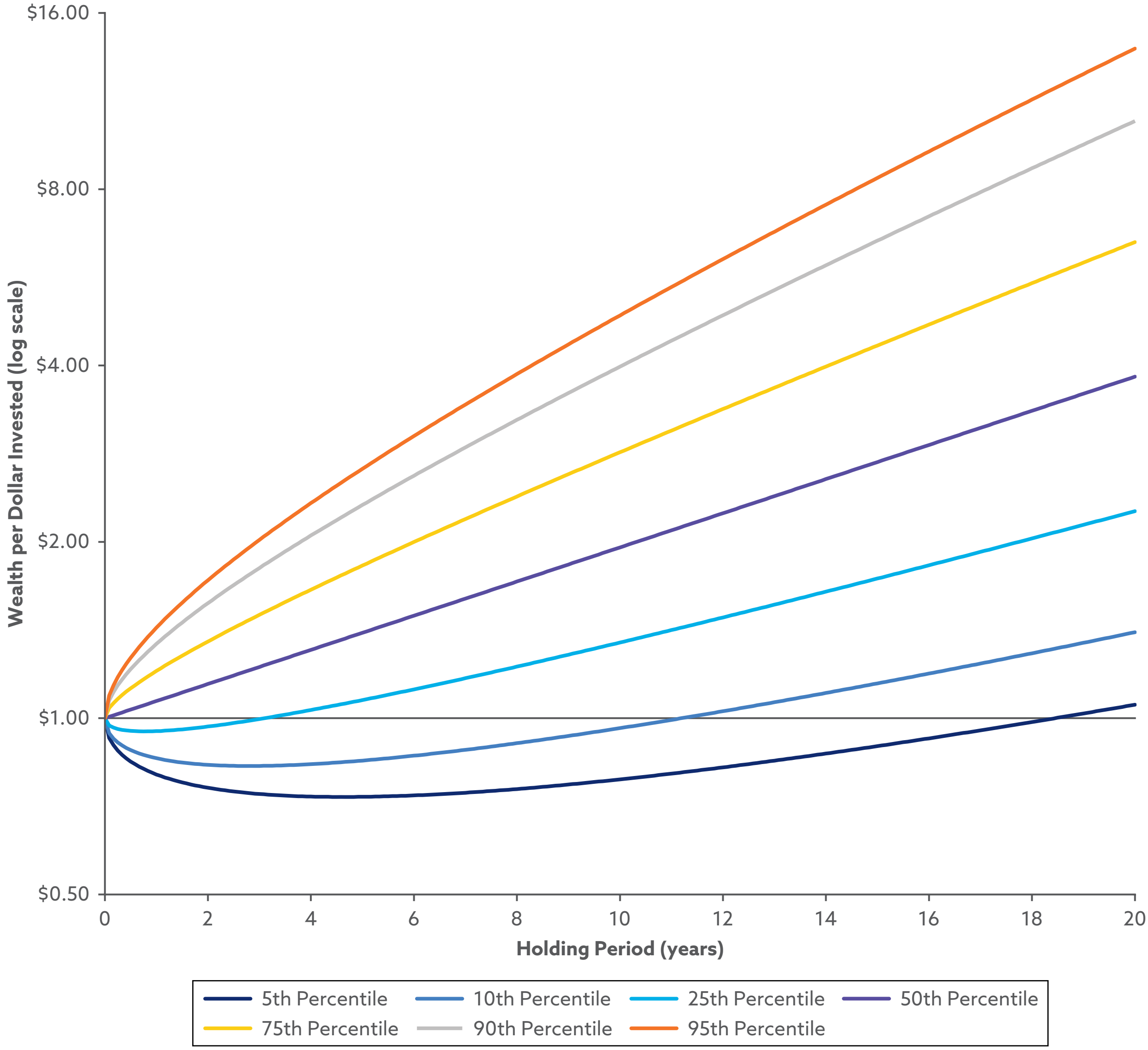
Exhibit 19.4 shows the percentiles of future wealth for large-cap stocks with W_0 = \$1 for the lognormal model. **Exhibit 19.5** shows percentiles for CAR. In these two exhibits, the horizontal axis depicts $T/12$, so it is in years.

Exhibit 19.3. Percentiles of the Standard Normal Distribution

Percentile	Standard Normal Value
5th	–1.6449
10th	–1.2816
25th	–0.6745
50th	0.0000
75th	0.6745
90th	1.2816
95th	1.6449

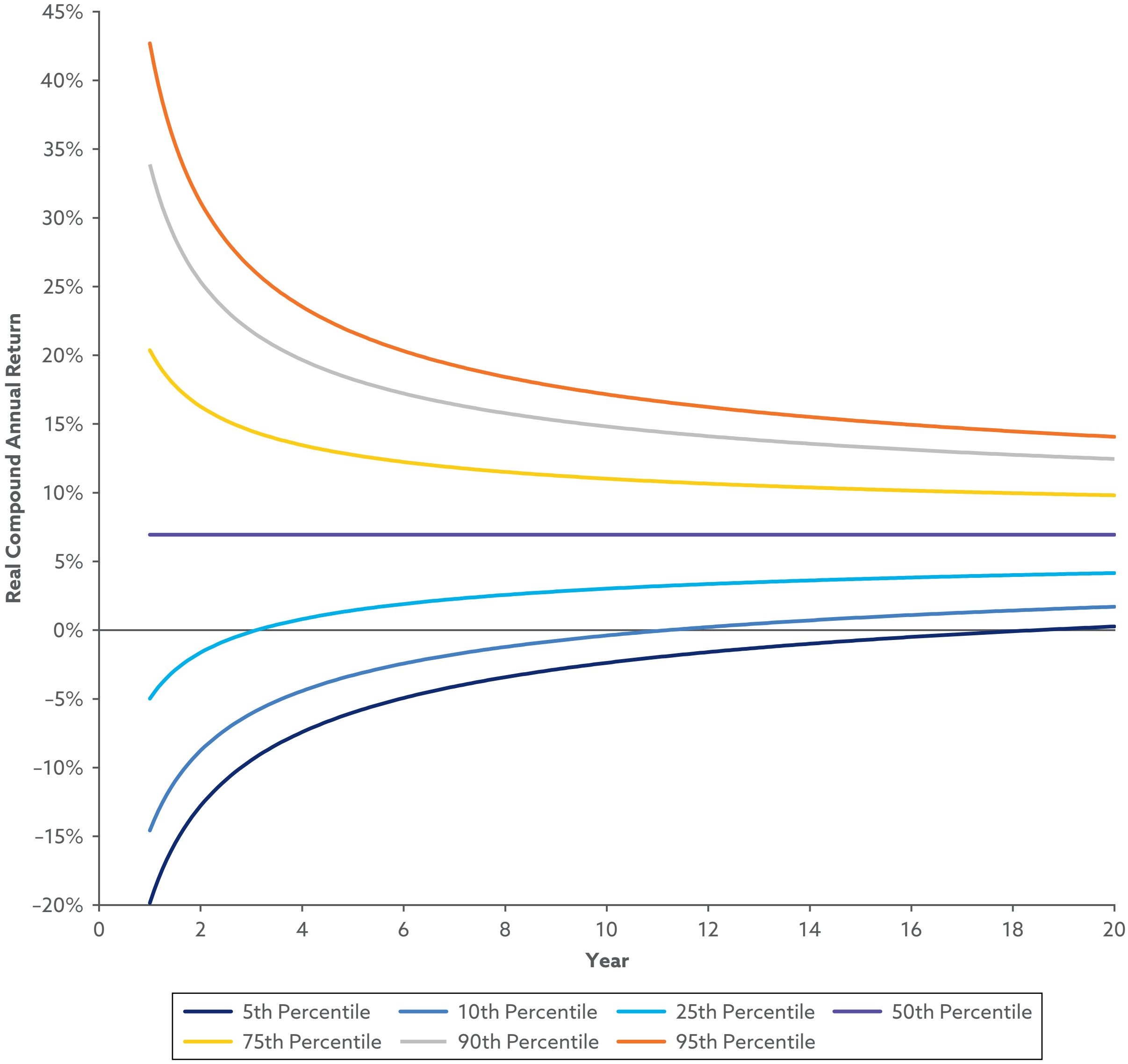
Source: Standard calculations.

Exhibit 19.4. Percentiles of Future Wealth Under the Lognormal Model: Large-Cap Stocks



Source: Calculations by the author using Ibbotson Indices data.

Exhibit 19.5. Percentiles of Future Compound Annual Return Under the Lognormal Model: Large-Cap Stocks



Source: Calculations by the author using Ibbotson Indices data.

The Central Limit Theorem and the Lognormal Distribution

Suppose that we have a sequence of T independent and identically distributed (i.i.d.) random variables, $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_T$, all with the same mean, μ, and all with the same standard deviation, σ.[6] These variables may or may not be normally distributed. The mean of this sequence is[7]

$$\tilde{\bar{X}}_T = \frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t. \qquad (19.S1)$$

Let

$$\tilde{Y}_T = (\tilde{\bar{X}}_T - \mu)\sqrt{T}. \qquad (19.S2)$$

$\tilde{Y}_T$ is the mean of the sequence, centered to have a mean of zero and scaled to have a standard deviation of σ. The central limit theorem says that as $t \to \infty$, $\tilde{Y}_T$ approaches a normally distributed random variable with a mean of zero and a standard deviation of σ. Hence, the mean and, therefore, the sum of a sequence of T i.i.d. random variables is well approximated by a normally distributed variable for large values of T.

To apply the central limit theorem to probabilistic forecasting of cumulative wealth, take logarithms of both sides of Equation 19.1:

$$\log(\tilde{W}_T) = \log(W_0) + \sum_{t=1}^{T}\log(1+\tilde{R}_{1t}). \qquad (19.S3)$$

Assuming that returns are i.i.d., applying the central limit theorem to Equation 19.S3, one can conclude that for large values of T, $\log(\tilde{W}_T)$ is well approximated by a normally distributed random variable with a mean of $\mu_1 T$ and a standard deviation of $\sigma_1\sqrt{T}$.

A random variable is lognormally distributed if its logarithm is normally distributed. Hence, because $\log(\tilde{W}_T)$ is well approximated by a normally distributed variable, it follows that $\tilde{W}_T$ is well approximated by a lognormally distributed random variable with logarithmic parameters $\mu_1 T$ and $\sigma_1\sqrt{T}$. It also follows that compound annual return, as defined by Equation 19.3, is well approximated by a lognormal distribution with logarithmic parameters μ_1 and $\sigma_1/\sqrt{T}$.

---

[6] That is, the realization of any one random variable in the sequence is completely unrelated to that of any other random variable in the sequence, and all the random variables in the sequence are drawn from the same distribution. A simple way to illustrate a sequence of i.i.d. random variables is to flip the same coin repeatedly. Note that the central limit theorem applies to any sequence of i.i.d. random variables, not just to time series.

[7] Following standard practice, I use a bar as part of the name of a variable to indicate that it is an average. Hence, the combination of a bar and a tilde indicate that the average itself is a random variable.

Restating this conclusion in plain English, when forecasting wealth at the end of T periods or (equivalently) the compound annual return over those T periods, the distribution can be assumed to be lognormal. Therefore, most such forecasts use the lognormal distribution. In this chapter, some of the forecasts use the lognormal distribution and others use an alternative distribution assumption called the Johnson distribution.

Although the lognormal distribution is considered to be heavy tailed, its tails decay too rapidly to be considered fat tailed.

The First Four Moments of Wealth Distributions

In the lognormal model, skewness and excess kurtosis are not specified by explicit parameters. Rather, these moments are functions of the logarithmic standard deviation given in Equation 19.5a.

To see how skewness and excess kurtosis relate to logarithmic standard deviation, consider the three-parameter lognormal distribution, which is a generalization of the lognormal distribution. If $\tilde{Z}$ is a standard normal random variable, $\tilde{X}$ follows a three-parameter lognormal distribution if[8]

$$\tilde{X} = \xi + \lambda \cdot \exp(\mu + \sigma \tilde{Z}). \tag{19.8}$$

In Equation 19.8, ξ is the third parameter of the three-parameter lognormal distribution. It shifts the distribution. The additional parameter, λ, determines whether the distribution is right skewed (+1) or left skewed (–1).

As discussed later, the three-parameter lognormal distribution is a limiting case of the Johnson distribution that will be used for parametric forecasting. For the remainder of this chapter, the term *lognormal distribution* will refer to the three-parameter lognormal distribution, with the two-parameter return distribution as a special case.

The major limitation of the lognormal model is that it does not explicitly take into account the shape of the short-term wealth distributions. (As discussed in the sidebar, according to the central limit theorem, the long-term distributions should be nearly lognormal.) This limitation can be seen from the formulas for the skewness and excess kurtosis of a lognormal distribution. Letting

$$\omega = \exp(\sigma^2), \tag{19.9}$$

the skewness is

$$s = \lambda(\omega + 2)\sqrt{\omega - 1} \tag{19.10a}$$

and the excess kurtosis is

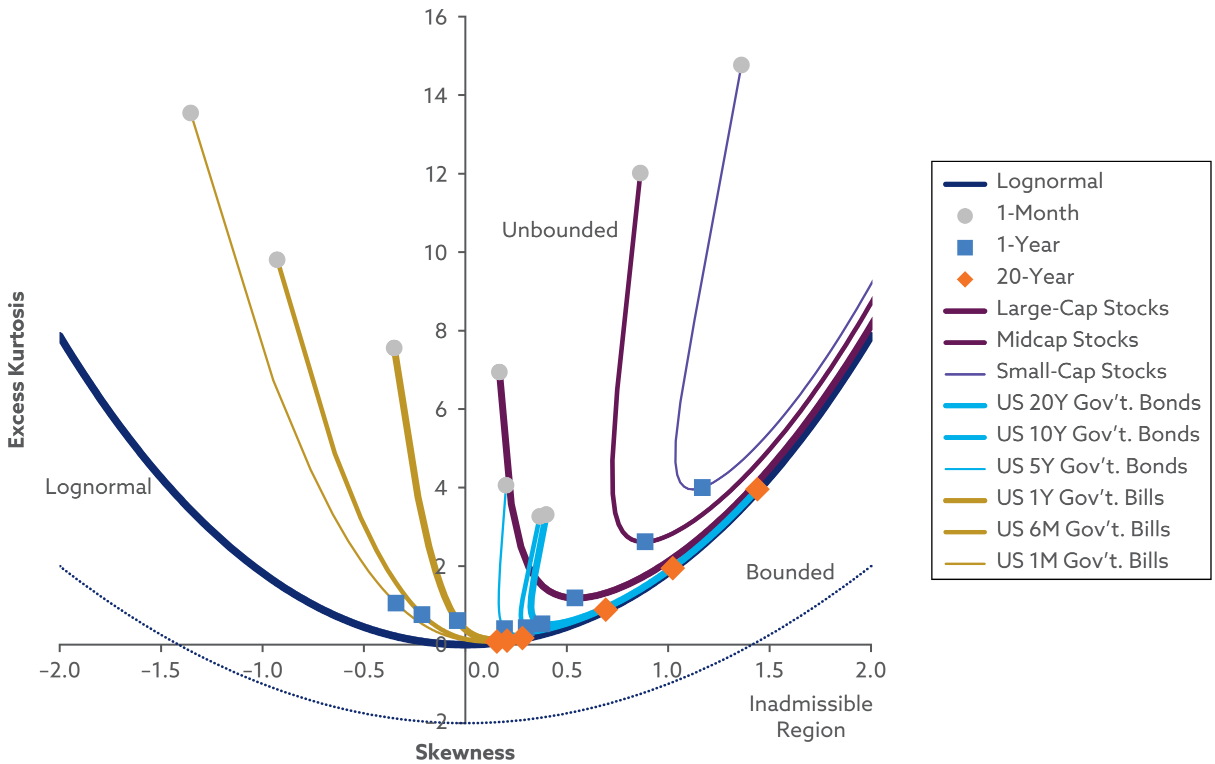
$$\kappa = \omega^4 + 2\omega^3 + 3\omega^2 - 6. \tag{19.10b}$$

[8] In equations that do not refer to a specific period of time, I omit the time period subscripts on the parameters.

The skewness and excess kurtosis of a lognormal distribution simultaneously satisfy Equations 19.10a and 19.10b. **Exhibit 19.6** shows possible combinations of skewness and excess kurtosis. The curve labeled "Lognormal" represents combinations for lognormal distributions implied by Equations 19.9 and 19.10. The gray dots depict the combinations for the asset class indices given in Exhibit 19.2. The asset classes are indicated by the color and thickness of the curves that emanate from the circles, as shown in the table just below the chart. Note that these combinations are far above the lognormal curve, indicating that monthly returns are far from lognormal. A different distribution, which I will introduce in the next section, is necessary.

In order to make probabilistic forecasts with Johnson distributions, the first four moments for any holding period need to be calculated. This calculation has two steps.

Exhibit 19.6. Combinations of Skewness and Excess Kurtosis for Stocks, Bonds, and Bills, and Regions for Johnson Distributions



	Color	Thick	Medium	Thin
Stocks	Purple	Large Cap	Mid Cap	Small Cap
US Treasury bonds	Light Blue	20 Years	10 Years	5 Years
US Treasury bills	Yellow	1 Year	6 Months	1 Month

Source: Calculations by the author using Ibbotson Indices data.

Step 1: Calculate the First Four Raw Monthly Moments

The *n*th monthly raw moment is defined as follows:

$$M_n = E[(1+\tilde{R}_1)^n]. \tag{19.11}$$

The raw moments from the moments are as follows:

$$M_1 = 1 + ER_1. \tag{19.12a}$$

$$M_2 = SD_1^2 + M_1^2. \tag{19.12b}$$

$$M_3 = s \cdot SD_1^3 + 3M_2M_1 - 2M_1^3. \tag{19.12c}$$

$$M_4 = (\kappa + 3)SD_1^4 + 4M_3M_1 - 6M_2M_1^2 + 3M_1^4. \tag{19.12d}$$

Step 2: Derive the First Four Moments for the Holding Period

If monthly returns are assumed to be independent over time and identically distributed, then

$$\mathrm{E}\left[\prod_{t=1}^{T}(1+\tilde{R}_{1t})^n\right] = M_n^T, \tag{19.13}$$

where T is the number of months of the holding period. Based on Equations 19.12 and 19.13, the first four moments for the holding period can be derived from the monthly raw moments as follows:

$$ER_T = M_1^T - 1. \tag{19.14a}$$

$$SD_T = \sqrt{M_2^T - M_1^{2T}}. \tag{19.14b}$$

$$s_T = \frac{M_3^T - 3M_2^TM_1^T + 2M_1^{3T}}{SD_T^3}. \tag{19.14c}$$

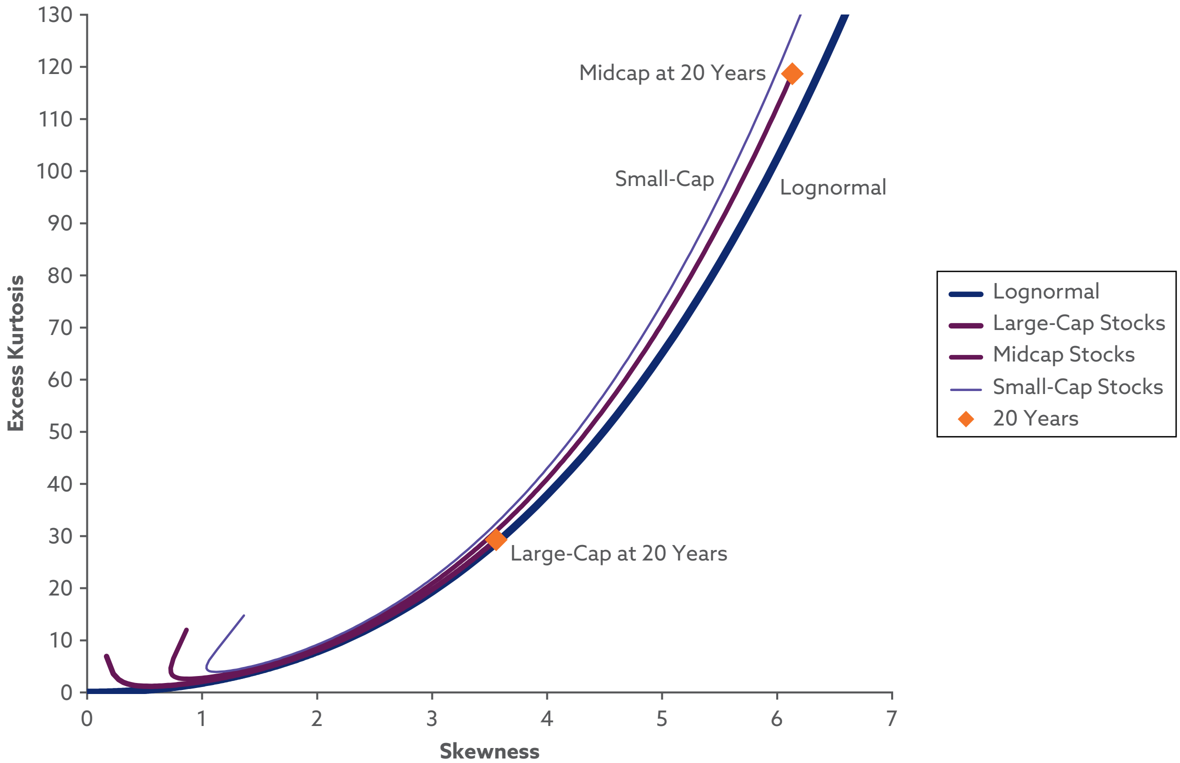
$$\kappa_T = \frac{M_4^T - 4M_3^TM_1^T + 6M_2^TM_1^{2T} - 3M_1^{4T}}{SD_T^4} - 3. \tag{19.14d}$$

Exhibit 19.6 shows the progression of combinations of skewness and excess kurtosis for the bond and bill asset class indices as the holding period is increased one month at a time. These curves start at 1 month (shown by the gray dots) and end at 240 months (20 years, shown by the orange diamonds). In each case, as the holding period increases, at first, the combination moves closer to the lognormal curve. Once the combination is near the curve, it moves in parallel to the curve. As discussed in the sidebar, this is a manifestation of the central limit theorem.

The ranges of values of skewness and excess kurtosis for the stock indices exceed the ranges of possible values in Exhibit 19.6, which is the reason only parts of the curves for these indices appear in this exhibit. To show the full progressions from 1 to 240 months of the combinations of skewness and excess kurtosis for large-cap and midcap stocks, in **Exhibit 19.7** I extend the range of possible values for skewness out to 7 and excess kurtosis out to 130. (This exhibit also shows part of the curve for small-cap stocks. The full curve is in **Exhibit 19.8**.) The curve for large-cap stocks comes very close to the lognormal curve, but the curve for midcap stocks is a bit above it, even at 20 years. In the next section, I will discuss the consequences of this finding for the Johnson forecasting model.

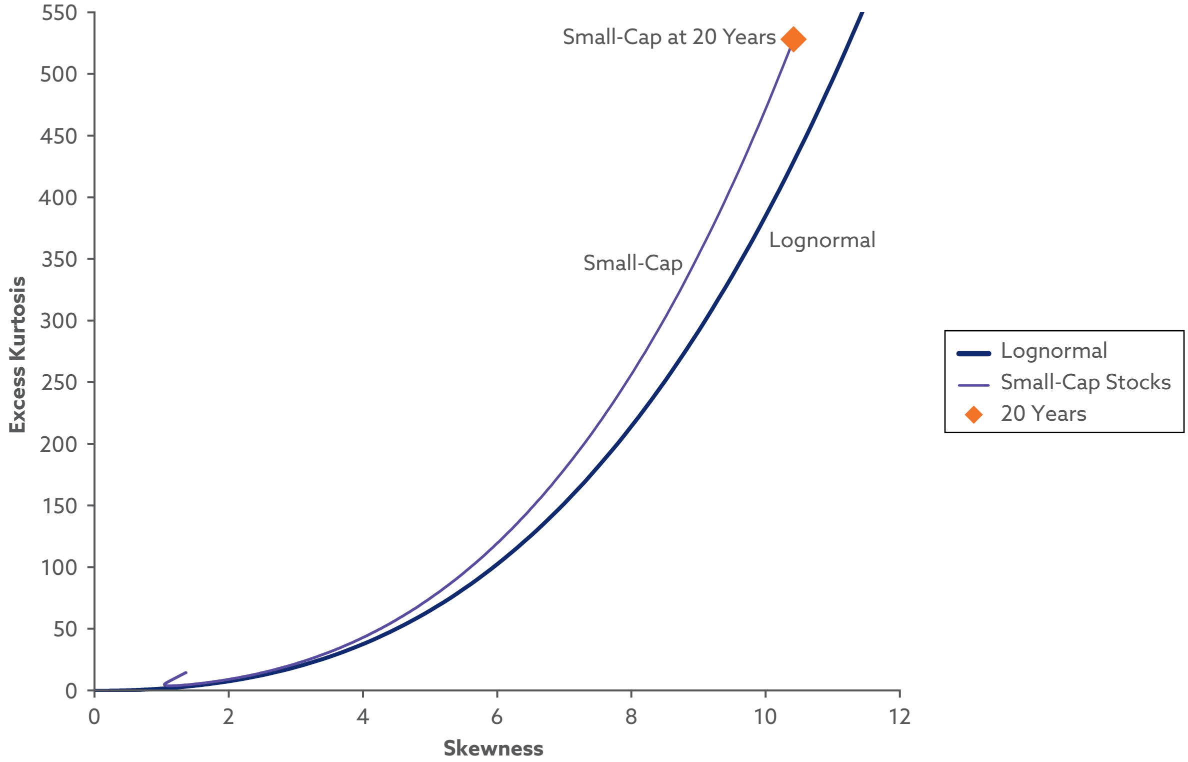
To show the full progression from 1 to 240 months of combinations of skewness and excess kurtosis curve for small-cap stocks, in Exhibit 19.8 I extend the range of possible values of skewness to 12 and excess kurtosis out to 550. The curve for small-cap stocks is considerably above the lognormal curve, even at 20 years, which has major repercussions in the Johnson forecasting model, discussed next.

Exhibit 19.7. Combinations of Skewness and Excess Kurtosis for Stocks



Source: Calculations by the author using Ibbotson Indices data.

Exhibit 19.8. Combinations of Skewness and Excess Kurtosis for Small-Cap Stocks



Source: Calculations by the author using Ibbotson Indices data.

The Johnson Forecasting Model

Johnson (1949) created a system of distributions based on combinations of skewness and excess kurtosis, depicted in Exhibit 19.6. First, not all combinations are possible. A combination is possible only if

$$\kappa \geq s^2 - 2. \tag{19.15}$$

The curve labeled "Admissible Region" in Exhibit 19.6 depicts Equation 19.15 as the dividing line between admissible and inadmissible combinations.

Among the admissible combinations, there are four cases:

- If $s = \kappa = 0$, the distribution is normal.
- If (s, κ) is on any other point on the lognormal curve, the distribution is lognormal.
- If (s, κ) is below the lognormal curve, the distribution is a bounded Johnson distribution.
- If (s, κ) is above the lognormal curve, the distribution is an unbounded Johnson distribution.

Because the combinations for all the asset class indices are above the lognormal curve, I will focus on the unbounded Johnson distribution. Thus, a four-parameter distribution is defined as follows for monthly returns:

$$1+\tilde{R}_1=\lambda_1 \sinh\left(\frac{\tilde{Z}-\gamma_1}{\delta_1}\right)+\xi_1, \tag{19.16}$$

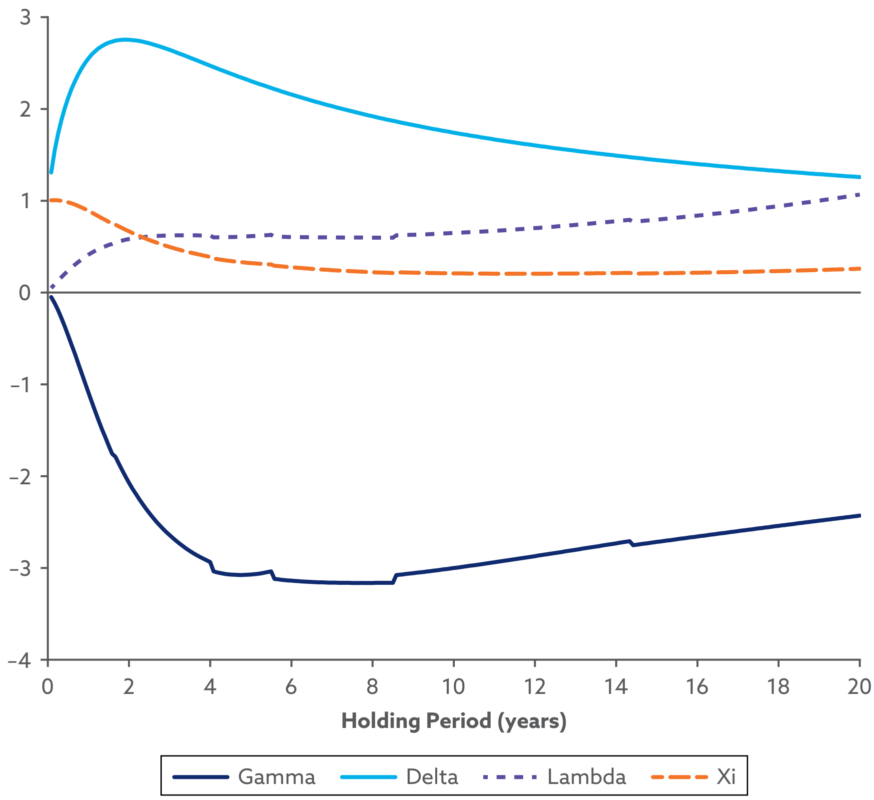
where sinh(·) is the hyperbolic sine function defined as

$$\sinh(x)=\frac{1}{2}[\exp(x)-\exp(-x)]. \tag{19.17}$$

Four parameters—λ_1, ξ_1, γ_1, δ_1—are calculated from the first four monthly moments using the algorithm developed by Hill, Hill, and Holder (1976).

To form probabilistic forecasts with the unbounded Johnson distribution, the four parameters are calculated for the Johnson distribution at T months from the first four moments given in Equation 19.14 using the Hill et al. (1976) algorithm. **Exhibit 19.9** shows the progression of the

Exhibit 19.9. Evolution of Parameters of the Johnson Distribution: Large-Cap Stocks



Source: Calculations by the author using Ibbotson Indices data.

parameters of the Johnson distribution for $T = 1, 2, \ldots, 240$ for large-cap stocks (the horizontal axis is in years, so it shows $T/12$).

From the parameters for a holding period of T months, the 100pth percentile of wealth is calculated as

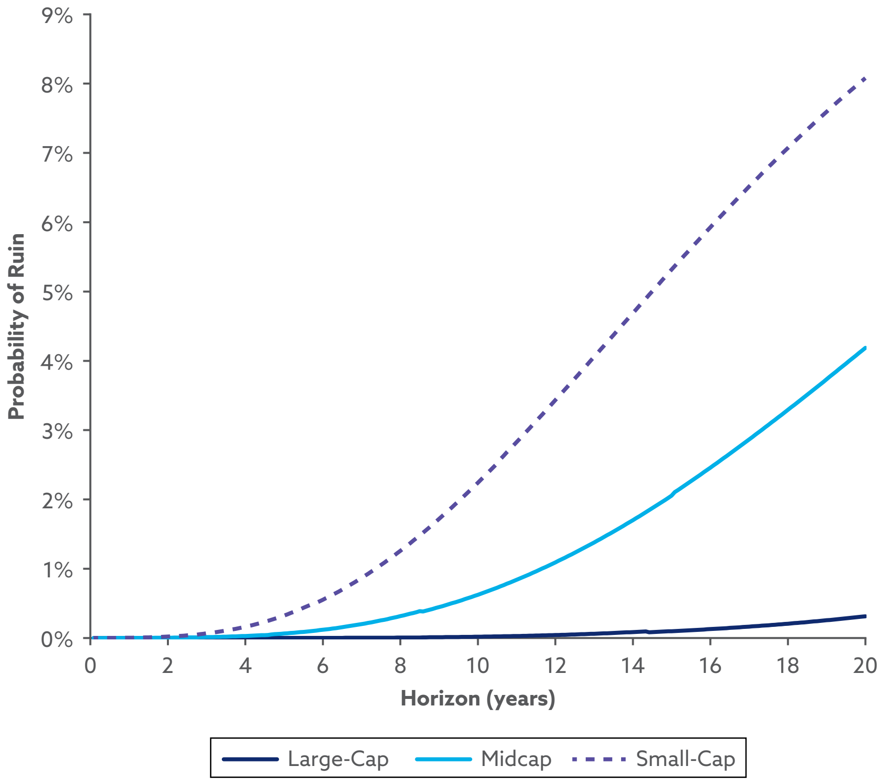
$$W_T(p) = W_0\left[\lambda_T \sinh\left(\frac{z(p)-\gamma_T}{\delta_T}\right)+\xi_T\right]. \tag{19.18}$$

Equation 19.18 can result in a negative value. Such an outcome can be interpreted to mean that a return of –100% is realized over the holding period, resulting in complete ruin; I define $\bar{p}(T)$ as the "probability of ruin" for the holding period of T months.[9] It is the value that satisfies

$$W_T\left[\bar{p}(T)\right] = 0. \tag{19.19}$$

Exhibit 19.10 shows the "probability of ruin" for large-cap, midcap, and small-cap stocks over holding periods from 1 to 240 months (20 years). **Exhibit 19.11** shows the "probability of

Exhibit 19.10. "Probability of Ruin" for Stocks: 1–240 Months



Source: Calculations by the author using Ibbotson Indices data.

[9] Later, I will explain why I put "probability of ruin" in quotes.

Exhibit 19.11. "Probability of Ruin" for Stocks at 240 Months and the Implied Frequency of Ruinous Monthly Returns

Asset Class	"Probability of Ruin"	Frequency of Ruinous Returns
Large-cap stocks	0.31%	1 month in 76,453 (6,371 years)
Midcap stocks	4.19%	1 month in 5,611 (468 years)
Small-cap stocks	8.08%	1 month in 2,849 (237 years)

Source: Ibbotson Indices. Calculations by the author.

ruin" for these stock indices at 240 months and the implied frequency of a monthly return of –100%.[10]

As Exhibit 19.10 shows, the "probability of ruin" increases with the holding period. The reason is that it takes only one ruinous return to wipe out all accumulated wealth. Exhibits 19.10 and 19.11 also show large differences in the "probability of ruin" between large-cap, midcap, and small-cap stocks. These differences result from the differences in the gaps between the paths of skewness and excess kurtosis for each asset class index and the lognormal curve in Exhibits 19.7 and 19.8.

For large-cap stocks, there is almost no gap; consequently, the "probability of ruin" is only 0.31% at 240 months. For midcap stocks, a small but noticeable gap exists, resulting in a "probability of ruin" of 4.19% at 240 months. For small-cap stocks, however, the gap is quite large, resulting in a "probability of ruin" of 8.08% at 240 months. Note that although a "probability of ruin" at 240 months of 8.08% may seem large, it implies that the frequency of ruinous returns is just one month in 237 years.

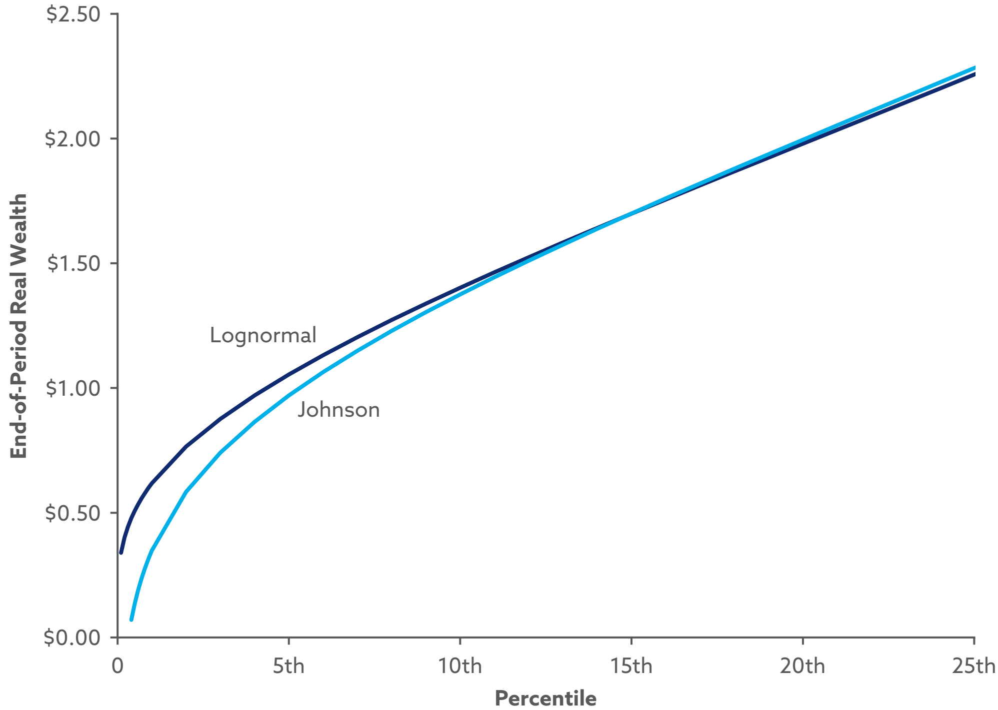
Of course, these values for the "probability of ruin" should not be taken literally. This is why I have placed the phrase "probability of ruin" in quotation marks. Although it might be possible for a single stock to lose 100% of its market value, such a loss is highly unlikely for a diversified portfolio of stocks. At each point in time, each stock index consists of many stocks, with the composition changing over time. In a given month, some stocks might lose 100% of their market value, but not all components of an index will do so at the same time (except in extreme circumstances, such as a revolution or losing a war).

What the values for the "probability of ruin" shown in Exhibit 19.10 and 19.11 do show is that the lognormal model can understate the riskiness of asset classes with return distributions that are far from being lognormal. The "probability of ruin" is not a forecast; rather, it is an output of the Johnson model that indicates the presence of risks not captured by the lognormal model.

The differences in the paths of skewness and excess kurtosis relative to the lognormal curve also result in corresponding differences between the wealth percentile curves from the lognormal and Johnson models. In the case of large-cap stocks, because the paths of skewness and kurtosis are very close to that of the lognormal, the results are also very close. Differences are

[10] The implied frequency of ruinous monthly returns is 1 in $\dfrac{1}{1-[1-\bar{p}(T)]^{\frac{1}{T}}}$ months.

Exhibit 19.12. Low Percentiles of Wealth at 20 Years for Large-Cap Stocks: Lognormal and Johnson Models



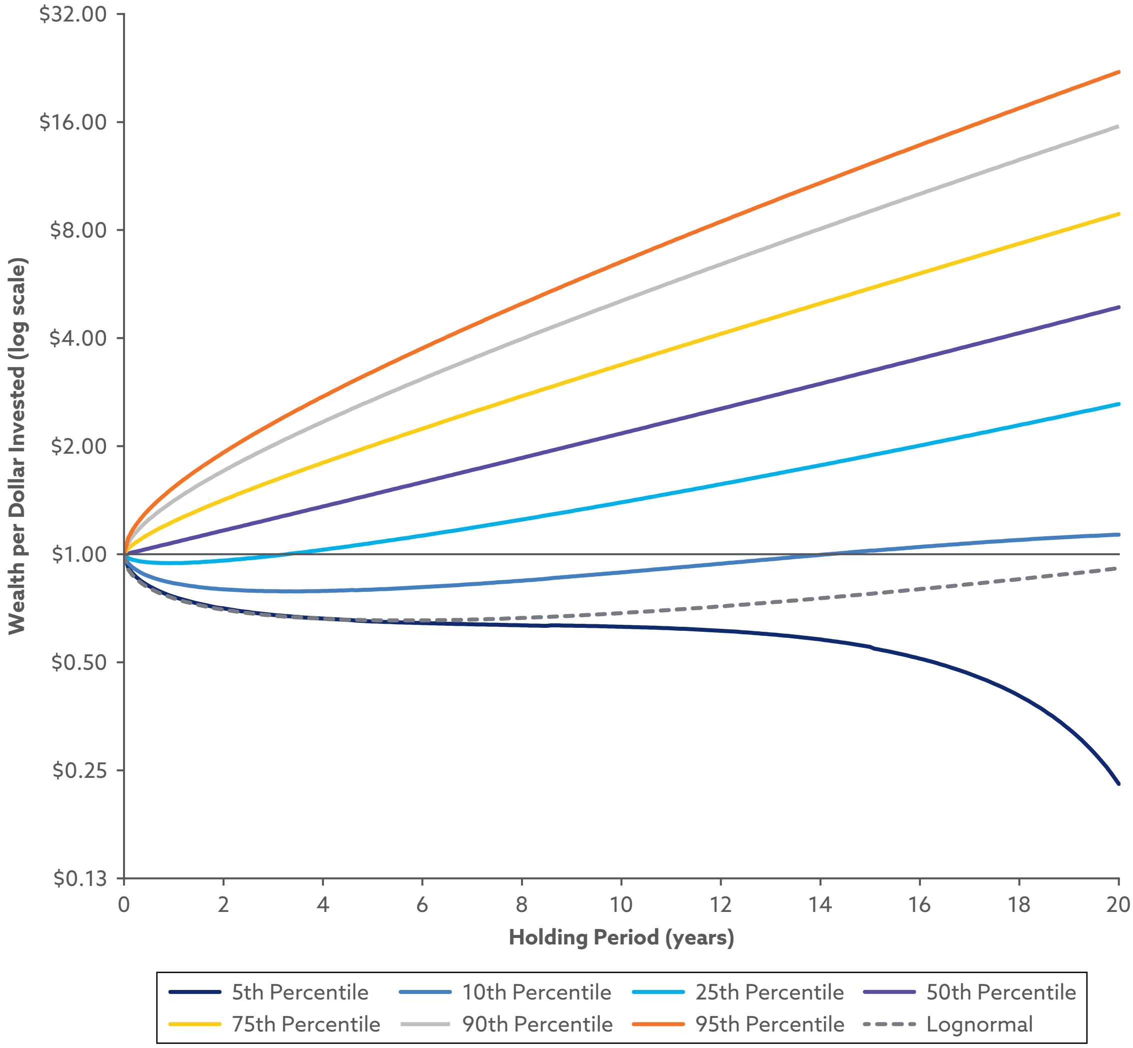
Source: Calculations by the author using Ibbotson Indices data.

apparent only for very low percentiles at long holding periods. To illustrate this, **Exhibit 19.12** shows percentiles of wealth in 20 years for both the lognormal and Johnson models, at percentiles varying from the 0.1th to the 25th. It is only at these extremes that the Johnson model is clearly more pessimistic than the lognormal model.

Exhibit 19.13 shows wealth percentiles for midcap stocks using the Johnson model. It also shows the 5th percentile using the lognormal model as a dashed curve. At the 5th percentile, the two models begin to deviate at a holding period of about six years, with the Johnson model always being more pessimistic. At about 12 years, while the lognormal curve continues to slope upward, the Johnson curve slopes downward, thus becoming even more pessimistic as the holding period increases. These differences between the 5th percentiles of wealth for the two models are further manifestations of the gap between the path of skewness and excess kurtosis for midcap stocks and the lognormal curve in Exhibit 19.7.

Exhibit 19.14 shows wealth percentiles for small-cap stocks using the Johnson model. It also shows the 5th and 10th percentiles using the lognormal model as burgundy dashed (5th) and gray dotted (10th) curves. At the 10th percentile, the two models begin to deviate at a holding period of about 12 years, with the Johnson model always being more pessimistic. While the lognormal curve continues to slope upward, the Johnson curve slopes downward, becoming even more pessimistic as the holding period increases.

Exhibit 19.13. Wealth Percentiles for Midcap Stocks Under the Johnson Model

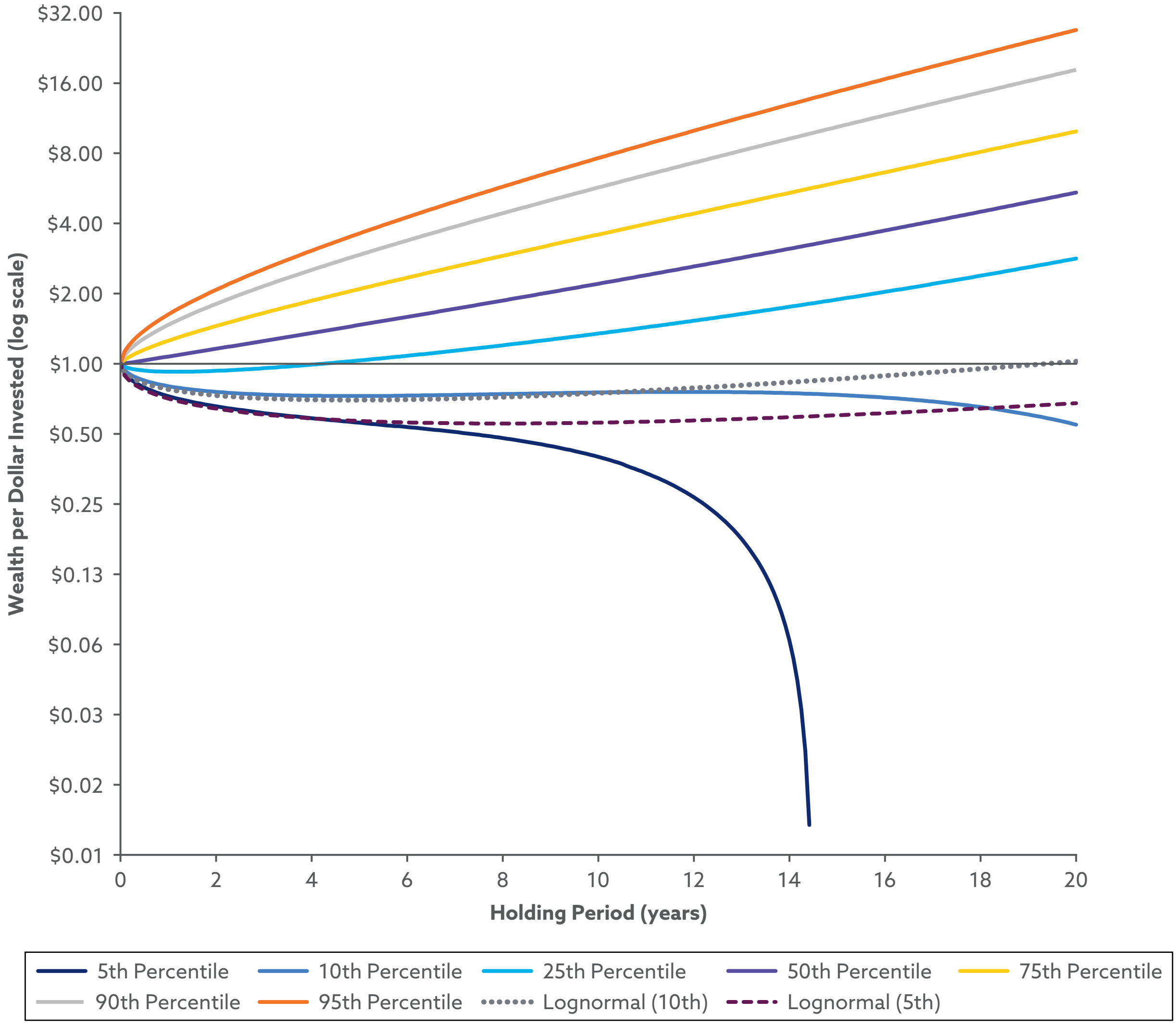


Source: Calculations by the author using Ibbotson Indices data.

The 5th percentile curve for the Johnson model begins to deviate from the lognormal model at about six years. It then drops—and at an increasing rate. It reaches $0 at 14.5 years because, as Exhibit 19.10 shows, this is where the "probability of ruin" for small-cap stocks reaches 5%.[11]

[11] Because the vertical axis of Exhibit 19.14 is logarithmic, wealth cannot actually be shown going to $0 in that exhibit. Instead, the curve ends at 14.5 years with a value of $0.01.

Exhibit 19.14. Wealth Percentiles for Small-Cap Stocks Under the Johnson Model



Source: Calculations by the author using Ibbotson Indices data.

These differences between the 10th and 5th percentiles of wealth for the lognormal and Johnson models result from the large gap between the paths of skewness and excess kurtosis for small-cap stocks and the lognormal curve in Exhibit 19.8.

Annualized Parameters and Applications

Asset class return models are typically stated in terms of annual returns. By setting $T = 12$ and applying Equations 19.11 and 19.12, the first four moments can be annualized. **Exhibit 19.15**

Exhibit 19.15. Annualized Estimates of the First Four Moments of Real Asset Class Indices, January 1926–December 2025

Asset Class	Mean	Standard Deviation	Skewness	Excess Kurtosis
Large-cap stocks	8.60%	19.19%	0.54	1.19
Midcap stocks	10.45%	24.36%	0.89	2.62
Small-cap stocks	11.21%	28.72%	1.17	4.00
US 20-year Treasury bonds	2.34%	9.67%	0.38	0.53
US 10-year Treasury bonds	2.23%	7.18%	0.30	0.43
US 5-year Treasury bonds	1.97%	5.03%	0.19	0.41
US 1-year Treasury bills	1.08%	2.29%	−0.04	0.61
US 6-month Treasury bills	0.82%	2.00%	−0.21	0.77
US 1-month Treasury bills	0.32%	1.83%	−0.34	1.06

Source: Calculations by the author using Ibbotson Indices data.

shows the results of doing this for the monthly statistics in Exhibit 19.2. Exhibit 19.6 showed how the annualized combinations of skewness and kurtosis are much closer to the lognormal curve.

Covariance and Correlation

Asset class return models need to model not only the distribution of each asset class but also the covariance of returns between every pair of asset classes. The monthly covariance of the returns of asset classes *j* and *k* is estimated as follows:

$$\hat{C}_{1jk} = \frac{1}{N-1}\sum_{i=1}^{N}(R_{1ji} - \widehat{ER}_{1j})(R_{1ki} - \widehat{ER}_{1k}). \tag{19.20}$$

Because covariances can be hard to interpret, correlations are often reported instead, because a correlation must be between −1 and +1. The monthly correlation of the returns of asset classes *j* and *k* is estimated as follows:

$$\hat{\rho}_{1jk} = \frac{\hat{C}_{1jk}}{\widehat{SD}_{1j}\,\widehat{SD}_{1k}}. \tag{19.21}$$

Exhibit 19.16 presents the estimated monthly correlations between all pairs of asset class indices.

Exhibit 19.16. Estimates of Monthly Correlations of Real Asset Class Indices, January 1926–December 2025

Asset Class	1	2	3	4	5	6	7	8	9
(1) Large-cap stocks	1.0000								
(2) Midcap stocks	0.9396	1.0000							
(3) Small-cap stocks	0.8749	0.9684	1.0000						
(4) US 20-year Treasury bonds	0.0948	0.0580	0.0213	1.0000					
(5) US 10-year Treasury bonds	0.1018	0.0694	0.0352	0.9342	1.0000				
(6) US 5-year Treasury bonds	0.0980	0.0692	0.0395	0.8341	0.9243	1.0000			
(7) US 1-year Treasury bills	0.1195	0.1032	0.0834	0.5442	0.6259	0.7561	1.0000		
(8) US 6-month Treasury bills	0.0910	0.0733	0.0578	0.4175	0.4825	0.6023	0.9509	1.0000	
(9) US 1-month Treasury bills	0.0741	0.0503	0.0397	0.2968	0.3477	0.4513	0.8476	0.9490	1.0000

Source: Calculations by the author using Ibbotson Indices data.

The formula for annualizing covariances is

$$\hat{C}_{12jk} = [\hat{C}_{1jk} + (1+\widehat{ER}_{1j})(1+\widehat{ER}_{1k})]^{12} - [(1+\widehat{ER}_{1j})(1+\widehat{ER}_{1k})]^{12}. \tag{19.22}$$

The annualized correlation of the returns of asset classes j and k is given by

$$\hat{\rho}_{12jk} = \frac{\hat{C}_{12jk}}{\widehat{SD}_{12j}\,\widehat{SD}_{12k}}. \tag{19.23}$$

Exhibit 19.17 presents the annualized correlations between all pairs of asset class indices. All the off-diagonal entries are slightly lower in magnitude than their monthly counterparts.

Exhibit 19.17. Annualized Correlations of Real Asset Class Indices, January 1926–December 2025

Asset Class	1	2	3	4	5	6	7	8	9
(1) Large-cap stocks	1.0000								
(2) Midcap stocks	0.9382	1.0000							
(3) Small-cap stocks	0.8714	0.9673	1.0000						
(4) US 20-year Treasury bonds	0.0940	0.0573	0.0209	1.0000					
(5) US 10-year Treasury bonds	0.1010	0.0686	0.0346	0.9339	1.0000				
(6) US 5-year Treasury bonds	0.0973	0.0685	0.0389	0.8334	0.9241	1.0000			
(7) US 1-year Treasury bills	0.1187	0.1021	0.0822	0.5433	0.6254	0.7596	1.0000		
(8) US 6-month Treasury bills	0.0904	0.0726	0.0569	0.4168	0.4821	0.5883	0.9508	1.0000	
(9) US 1-month Treasury bills	0.0736	0.0497	0.0391	0.2962	0.3473	0.4397	0.8475	0.9490	1.0000

Source: Calculations by the author using Ibbotson Indices data.

Capital Market Assumptions for Mean–Variance Optimization

The annualized asset class expected returns, standard deviations, and correlations in Exhibits 19.15 and 19.17 together form a set of capital market assumptions (CMAs) that are the inputs to Markowitz (1952, 1959, 1987) mean–variance optimization (MVO). It is not good practice, however, to perform MVO with CMAs, especially expected returns, that are based purely on historical data, because doing so amounts to assuming that the future will be just like the past.

Idzorek and Kaplan (2024) developed a method for creating a set of expected returns that are calibrated to a complete contingent claims market model. Financial economists often use this type of approach to model capital markets. In this type of model, there is a single random variable—called the stochastic discount factor (SDF), sometimes called a "pricing kernel"—for pricing all securities. Kaplan and Idzorek (2025) explained their approach in more detail. The following discussion is based on Kaplan and Idzorek (2025).

Reverse Optimization

I begin the process of setting expected returns with the reverse optimization process devised by Sharpe (1974). Reverse optimization, as the name implies, is mean–variance optimization run in reverse.

In the standard *forward* mean–variance optimization model, portfolios of discrete investments or mixes of asset classes are constructed to minimize portfolio variance for a given level of portfolio expected return. The model requires two inputs:

$\mathbf{E_{AC}}$ = a vector, or list, of K asset class expected returns.

$\mathbf{V_{AC}}$ = a $K \times K$ covariance matrix of asset class returns.

Note that in writing formulas, *italic* font is used for scalars and **bold** font is used for vectors and matrices.

E_p is used to denote the target expected return for the portfolio, and $\mathbf{h}$ is used to denote the vector of portfolio weights, or holdings. The MVO model solves the following constrained quadratic program:

$$V_{MVO}(E_P) = \min_{\mathbf{h}} \mathbf{h}^{\mathrm{T}} \mathbf{V_{AC}} \mathbf{h} \text{ s.t. } \mathbf{h}^{\mathrm{T}} \mathbf{E_{AC}} = E_P, \sum_{i=1}^{K} h_i = 1, \mathbf{h} \geq 0. \tag{19.24}$$

Equation 19.24 states that the optimization procedure finds the portfolio weights (**h**) for the K asset classes that *minimize* portfolio variance subject to (s.t.) three constraints: (1) The portfolio return must equal E_p, (2) the weights of the portfolio must sum to 1—that is, 100%—and (3) each weight must be zero or greater.

Forward optimization begins with a set of *expected returns* ($\mathbf{E_{AC}}$) and arrives at efficient *weights* (**h**). By varying μ_p from the highest asset class expected return to the expected return of the minimum-variance portfolio, the Markowitz efficient frontier can be traced out.

In *reverse* optimization, one starts with a single set of *weights* (**h**) from a reference portfolio ($\mathbf{h_R}$) presumed to be efficient, such as a market-capitalization-weighted portfolio, and calculates the set of *expected returns* ($\boldsymbol{\mu}_{\mathbf{AC}}$) that make those weights efficient. With this construct, a closed-form solution exists, and a quadratic program is unnecessary. For the reference portfolio to be on the efficient frontier, the following condition must be true:

$$\mathbf{E_{AC}} = r_f + (E_R - r_f)\boldsymbol{\beta}, \tag{19.25}$$

where r_f is the expected return on a portfolio with a beta of zero. (This expected return is denoted r_f because it is set to the risk-free rate.) E_R is the expected return of the reference portfolio discussed below, and $\boldsymbol{\beta}$ is the vector of asset class betas (defined below) with respect to the reference portfolio.

Similar to the capital asset pricing model (CAPM) formula for expected returns, Equation 19.25 states that the expected total return of an asset class equals the risk-free rate plus a risk premium multiplied by the beta of the asset class relative to the reference portfolio.[12] The most common reference portfolio is a market-cap-weighted portfolio. The vector of asset class betas is given by

$$\beta = \frac{1}{\mathbf{h}_R^{\mathrm{T}} \mathbf{V}_{\mathbf{AC}} \mathbf{h}_{\mathbf{R}}} \mathbf{V}_{\mathbf{AC}} \mathbf{h}_{\mathbf{R}}. \tag{19.26}$$

In Equation 19.26, each asset class beta is calculated by dividing the covariance of each asset class, calculated relative to the reference portfolio, by the variance of the reference portfolio.

The expected return of the asset classes ($\mathbf{E_{AC}}$) can then be written as a function of μ_R and r_f as follows:

$$\mathbf{E_{AC}}(E_R, r_f) = r_f + (E_R - r_f)\boldsymbol{\beta}. \tag{19.27}$$

Based on Equation 19.27, **Exhibit 19.18** shows

- the reference portfolio that is assumed for this example;
- the betas computed from the reference portfolio, the standard deviations in Exhibit 19.15, and the correlations in Exhibit 19.17; and
- the expected returns on the asset classes using Equation 19.26, with r_f = 2% and E_R = 4.57%.

The next section explains how we set E_R by calibrating the expected returns.

Calibration of Expected Returns

Now E_R and r_f can be included as parameters that provide the expected returns on the asset classes ($\mathbf{E_{AC}}$) in an efficient frontier function:[13]

$$V_{MVO}(E_P; E_R, r_f) = \min_{\mathbf{h}} \mathbf{h}^{\mathrm{T}} \mathbf{V_{AC}} \mathbf{h} \,\text{s.t.}\, \mathbf{h}^{\mathrm{T}} \mu_{\mathbf{AC}}(E_R, r_f) = E_P, \sum_{i=1}^{K} h_i = 1,\ \mathbf{h} \geq 0. \tag{19.28}$$

Equation 19.28 is very similar to Equation 19.24, except that it has two additional parameters—the assumed expected return on the reference portfolio (E_R) and the risk-free rate (r_f). By varying across the range of possible portfolios returns (E_P), Equation 19.28 leads to an entire efficient frontier.

I chose to write many of the equations as functions of other values to clarify and automate the eventual calibration process, in which one altered value cascades across the various formulas. In this case, by changing the *assumed* expected total return on the reference portfolio, E_R, or the risk-free rate, r_f, or both, one arrives at a new set of expected returns ($\mathbf{E_{AC}}$) and, ultimately, a new

[12] In the CAPM, the reference portfolio is the market portfolio. Reverse optimization is more general than the CAPM in that the reference portfolio can be any portfolio assumed to be on the efficient frontier.

[13] In Markowitz (1952, 1959, 1987), the efficient frontier is a plot of the efficient frontier function as defined in Equation 19.28. Common practice, however, is to reverse the axes and plot the standard deviation rather than the variance of return. I follow the common practice in this chapter. See Exhibit 19.19.

Exhibit 19.18. Betas and Expected Returns from Reverse Optimization

Asset Class	Equity Portfolio	Reference Portfolio	Beta	Expected Return	Standard Deviation
Large-cap stocks	70.00%	42.00%	1.43	5.68%	19.19%
Midcap stocks	20.00%	12.00%	1.78	6.58%	24.36%
Small-cap stocks	10.00%	6.00%	1.99	7.10%	28.72%
US 20-year Treasury bonds	0.00%	13.33%	0.21	2.55%	9.67%
US 10-year Treasury bonds	0.00%	13.33%	0.16	2.42%	7.18%
US 5-year Treasury bonds	0.00%	13.33%	0.11	2.28%	5.03%
US 1-year Treasury bills	0.00%	0.00%	0.04	2.11%	2.29%
US 6-month Treasury bills	0.00%	0.00%	0.03	2.07%	2.00%
US 1-month Treasury bills	0.00%	0.00%	0.02	2.05%	1.83%
Equity portfolio	100.00%	—	1.56	6.00%	20.72%
Reference portfolio	—	100.00%	1.00	4.56%	12.98%

Source: Calculations by the author using Ibbotson Indices data.

efficient frontier. In practice, only the expected total return on the reference portfolio is varied and the risk-free rate is taken as a given.

To create consistency between the utility functions in many economic models and the MVO model, the mean–variance problem is set up using the mean–variance approximation of expected utility proposed by Levy and Markowitz (1979).[14] We assume a constant relative risk aversion (CRRA) utility function with risk tolerance parameter θ.

Based on Equation 19.28, it is possible to find the point on the efficient frontier that maximizes the Levy–Markowitz CRRA utility function for given values of θ, E_R, and r_f as follows:

$$\mu_{\theta}^{MVO}(E_R, r_f) = \arg\max_{\mu_P} U_{\theta}[E_P, V_{MVO}(E_P; E_R, r_f)], \qquad (19.29)$$

where $U_{\theta}(\cdot,\cdot)$ is the Levy–Markowitz CRRA utility function with risk tolerance θ described by Kaplan (2020), Idzorek and Kaplan (2024), Kaplan and Idzorek (2025), and Kaplan (2025).

[14] See Kaplan (2020), Idzorek and Kaplan (2024), Kaplan and Idzorek (2025), and Kaplan (2025).

The standard deviation at the point on the efficient frontier that maximizes the Levy–Markowitz CRRA utility function is

$$S_\theta^{MVO}(E_R,r_f)=\sqrt{V_{MVO}[E_\theta^{MVO}(E_R,r_f);E_R,r_f]}. \tag{19.30}$$

Kaplan (2025) explained how to use *parametric* quadratic programming to find the corner portfolios of the efficient frontier, identify them using the Levy–Markowitz utility function, and create the entire efficient frontier.

Inputting reverse optimized expected returns into a mean–variance optimizer and maximizing the Levy–Markowitz CRRA utility function, with various values of the risk tolerance parameter, θ, allows the calibration of the CMAs to the implied CMAs embedded in the complete contingent claims market model.

To do so, two sets of points are formed for a set of n values of θ in expected return–standard deviation space—one from the complete contingent claims market of the life-cycle model described in Kaplan and Idzorek (2025, Appendix A) and one from a mean–variance-efficient frontier based on the Levy–Markowitz CRRA utility function. Taking r_f as given, E_R is selected to minimize the sum of the squared Euclidean distances between the n corresponding points from each model, as follows:

$$E_R^*=\arg\min_{E_R}\sum_{j=1}^{n}\left[E_{\theta_j}^{SDF}-E_{\theta_j}^{MVO}(E_R,r_f)\right]^2+\left[S_{\theta_j}^{SDF}-S_{\theta_j}^{MVO}(E_R,r_f)\right]^2, \tag{19.31}$$

where

$\theta_1, \theta_2, \ldots, \theta_n$ = the n values of θ

$E_{\theta_j}^{SDF}$ = the expected return from the complete contingent claims market model with risk tolerance θ_j

$S_{\theta_j}^{SDF}$ = the standard deviation of return from the complete contingent claims market model with risk tolerance θ_j

To calculate $E_{\theta_j}^{SDF}$ and $S_{\theta_j}^{SDF}$, start with the definition of the SDF. The SDF is a random variable $\tilde{Q}$ such that the market price of any security with a random payout $\tilde{X}$ is $E[\tilde{Q}\tilde{X}]$. Because a risk-free asset always pays \$1,[15]

$$E[\tilde{Q}]=\frac{1}{1+r_f}. \tag{19.32}$$

The ideal portfolio of an investor with risk tolerance θ has a return of

$$\tilde{R}_\theta=\frac{\tilde{Q}^{-\theta}}{E[\tilde{Q}^{1-\theta}]}-1. \tag{19.33}$$

[15] A security that costs $\frac{1}{1+r_f}$ today and pays \$1 tomorrow with certainty is paying a riskless interest rate of r_f.

The SDF is assumed to follow a lognormal distribution. Hence,

$$\log(\tilde{Q}) \sim N\left[-\log(1+r_f)-\frac{1}{2}\sigma_{SDF}^2, \sigma_{SDF}^2\right], \tag{19.34}$$

where σ_{SDF}^2 is the logarithmic variance of the SDF. This parameter is to be determined.

To calculate the expected return and the standard deviation of the return given by Equation 19.33, first, the price of a security that pays $\tilde{Q}^{-\theta}$ needs to be calculated:

$$P_\theta = \mathrm{E}[\tilde{Q}^{1-\theta}] = (1+r_f)^{\theta-1}\exp\left[\frac{1}{2}\theta(\theta-1)\sigma_{SDF}^2\right]. \tag{19.35}$$

The expected value of $\log(1+\tilde{R}_\theta)$ is

$$\mathrm{E}[\log(1+\tilde{R}_\theta)] = \theta\left[\log(1+r_f)+\frac{1}{2}\sigma_{SDF}^2\right]-\log(P_\theta). \tag{19.36}$$

The expected return is

$$E_\theta^{SDF} = \exp\left\{\mathrm{E}[\log(1+\tilde{R}_\theta)]+\frac{1}{2}\theta^2\sigma_{SDF}^2\right\}-1. \tag{19.37}$$

The standard deviation of return is

$$S_\theta^{SDF} = (1+E_\theta^{SDF})\sqrt{\exp(\theta^2\sigma_{SDF}^2)-1}. \tag{19.38}$$

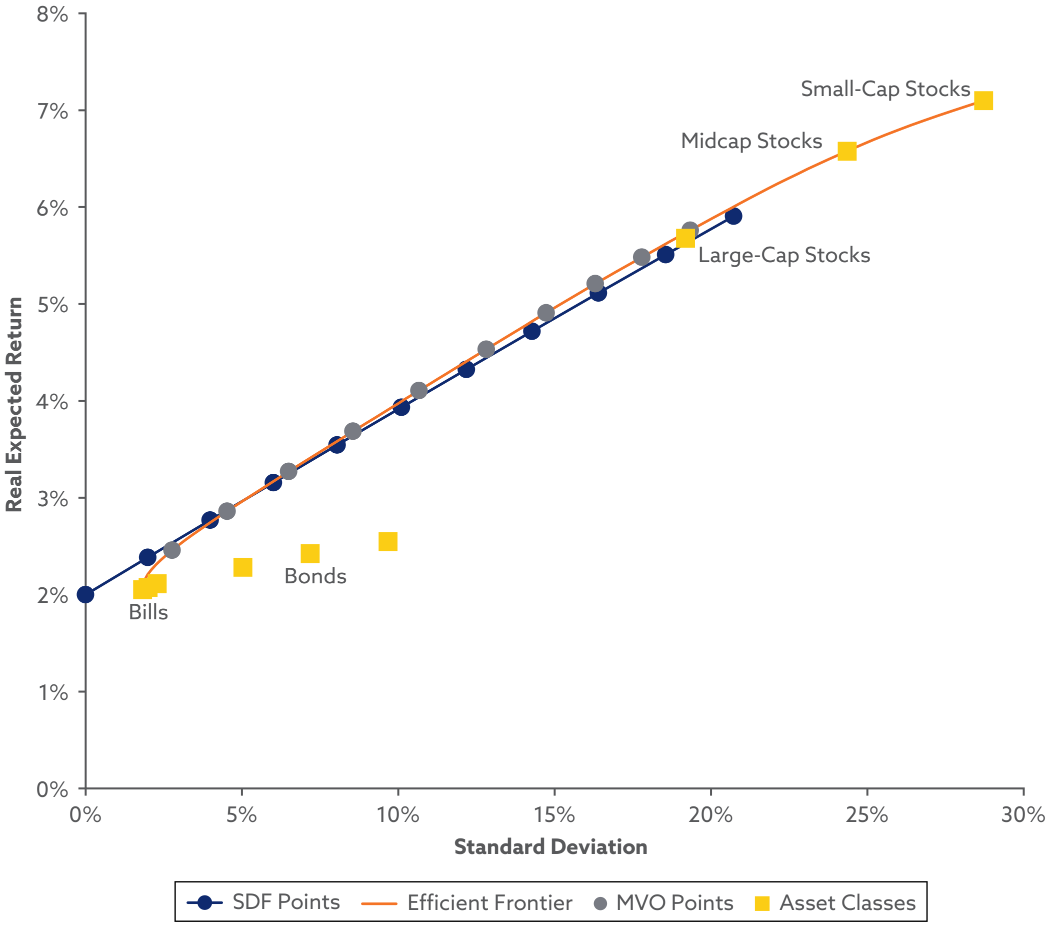
To find σ_{SDF}^2, it is assumed that the portfolio with θ = 100% is the equity portfolio in Exhibit 19.18 (this is the normalized equity subportfolio of the reference portfolio). The annualized covariance matrix shows that the standard deviation of this portfolio is 20.72%. $S_{100\%}^{SDF}$ is equated to this value. Using Equations 19.35–19.38,

$$\left(\frac{S_{100\%}^{SDF}}{1+r_f}\right)^2 = [\exp(\sigma_{SDF}^2)]^2[\exp(\sigma_{SDF}^2)-1]. \tag{19.39}$$

Solving Equation 19.39 for σ_{SDF} results in σ_{SDF} = 19.38%.

Exhibit 19.19, adopted from Idzorek and Kaplan (2024) and Kaplan and Idzorek (2025), illustrates the calibration process. Starting at the bottom, the connected blue dots show the 10 risk–expected return combinations associated with 10 different values for θ — 10%, 20%,…, 100% — coming from Equations 19.37 and 19.38. The gray dots on the orange efficient frontier identify 10 risk–expected return combinations from maximizing the Levy–Markowitz CRRA utility function for the same 10 values of θ. Equation 19.31 is solved for E_R^*, the expected return that minimizes the sum of the squared Euclidean distances between the 10 corresponding risk–expected return points from each model. (The 10 values are used to span the relevant part of the efficient frontier without making the minimization process too computationally intense.) The orange curve is simply the complete mean–variance-efficient frontier as defined in Equation 19.28.

Exhibit 19.19. Calibration Process for Expected Returns in Mean–Variance Optimization



Source: Calculations by the author using Ibbotson Indices data.

Key Takeaways

- Parametric forecasting models allow the formation of probabilistic forecasts of cumulative wealth and compound annual returns from moments of a return distribution. These forecasts consist of percentiles of wealth or return at various holding periods.
- The lognormal forecasting model uses only mean and standard deviation because the skewness and excess kurtosis of a lognormal distribution depend solely on its mean and standard deviation.
- The lognormal forecasting model works well for well-diversified portfolios and is the model of choice in most situations.

- If at long horizons the skewness and excess kurtosis of a wealth distribution do not converge to those of a lognormal distribution, the lognormal forecasting model is not appropriate.
- The Johnson forecasting model is an alternative to the lognormal forecasting model that explicitly takes skewness and kurtosis into account.
- In extreme cases, such as wealth distributions for small-cap stocks, the Johnson forecasting model results in distributions with large left tails at long horizons.
- Although the standard deviations and correlations used as inputs for mean–variance optimization are usually estimated using historical return data, it is important that the expected returns be consistent with an economic model. I present a way for doing this that uses reverse optimization.

References

Bollerslev, Tim. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31 (3): 307–27. doi:10.1016/0304-4076(86)90063-1.

Hill, I. D., R. Hill, and R. L. Holder. 1976. "Algorithm AS 99: Fitting Johnson Curves by Moments." *Applied Statistics* 25 (2): 180–89. doi:10.2307/2346692.

Ibbotson, Roger G., and Rex A. Sinquefield. 1976a. "Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926–1974)." *Journal of Business* 49 (1): 11–47. doi:10.1086/295803.

———. 1976b. "Stocks, Bonds, Bills, and Inflation: Simulations of the Future (1976–2000)." *Journal of Business* 49 (3): 313–38. doi:10.1086/295854.

Idzorek, Thomas M., and Paul D. Kaplan. 2024. *Lifetime Financial Advice: A Personalized Optimal Multilevel Approach*. Charlottesville, VA: CFA Institute Research Foundation. https://rpc.cfainstitute.org/research/foundation/2024/lifetime-financial-advice-a-personalized-optimal-multilevel-approach.

Johnson, N. L. 1949. "Systems of Frequency Curves Generated by Methods of Translation." *Biometrika* 36 (1/2): 149–76. doi:10.2307/2332539.

Kaplan, Paul D. 2012a. "Déjà Vu All Over Again." In *Frontiers of Modern Asset Allocation*, edited by Paul D. Kaplan, 211–22. Wiley. doi:10.1002/9781119205401.ch18.

———. 2012b. "Déjà Vu Around the World." In *Frontiers of Modern Asset Allocation*, edited by Paul D. Kaplan, 223–37. Wiley. doi:10.1002/9781119205401.ch19.

———. 2020. "Solving the Asset Location Problem, Part II: How to Expand Mean–Variance Optimization to Build a Tax-Aware Model." *Morningstar Magazine* (Q4): 64–67. https://magazine.morningstar.com/issues/q4-2020/solving-the-asset-location-problem-part-ii.

———. 2025. "Solving the Net Worth Optimization Problem." *Financial Planning Review* 8 (4). https://doi.org/10.1002/cfp2.70019.

Kaplan, Paul D., and Thomas M. Idzorek. 2025. "A Hybrid Lifecycle Net Worth Optimization Model." *Financial Planning Review* 8 (4). https://doi.org/10.1002/cfp2.70018.

Levy, Haim, and Harry M. Markowitz. 1979. "Approximating Expected Utility by a Function of Mean and Variance." *American Economic Review* 69 (3): 308–17. www.jstor.org/stable/1807366.

Lewis, Alan L., Sheen T. Kassouf, R. Dennis Brehm, and Jack Johnston. 1980. "The Ibbotson–Sinquefield Simulation Made Easy." *Journal of Business* 53 (2): 205–14. doi:10.1086/296081.

Markowitz, Harry M. 1952. "Portfolio Selection." *Journal of Finance* 7 (1): 77–91. doi:10.1111/j.1540-6261.1952.tb01525.x.

———. 1959. *Portfolio Selection: Efficient Diversification of Investments*. Hoboken, NJ: John Wiley & Sons.

———. 1987. *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Basil Blackwell.

Sharpe, William F. 1974. "Imputing Expected Security Returns from Portfolio Composition." *Journal of Financial and Quantitative Analysis* 9 (3): 463–72. doi:10.2307/2329873.

CHAPTER 20. FORECASTING EQUITY AND BOND RETURNS

Roger G. Ibbotson, PhD
Professor in the Practice Emeritus of Finance, Yale School of Management
Chair, Zebra Capital Management, LLC

William N. Goetzmann, PhD
Edwin J. Beinecke Professor of Finance and Management Studies
Faculty Director, International Center for Finance, Yale School of Management

Otto S. Manninen
Research Associate, Yale School of Management

Bootstrapping Future Return Distributions

In this chapter, we revisit the landmark forecasting methodology of Ibbotson and Sinquefield (1976b), bringing it forward with modern bootstrapping techniques and comparisons drawn from the five decades of research that followed that original work. Our goal is practical: to produce credible long-horizon return forecasts that are grounded in history, disciplined by current market conditions, and stress tested against the widest available international evidence.

The approach belongs to a family of forecasting methods that are simpler in spirit than the parametric models covered in the previous chapter—and often just as useful. Rather than fitting an explicit statistical model to return data, these methods construct future scenarios by resampling directly from the historical record. The past, rearranged in new sequences, becomes the raw material for the future. That is the core idea of bootstrapping, and it is a powerful one.

Bootstrap methods are particularly well suited to financial data. Asset returns are fat-tailed and skewed—properties that parametric models can struggle to capture cleanly. By drawing directly from observed historical returns, bootstrap forecasts pick up these features automatically, without needing to specify or defend a particular distributional assumption.

The Ibbotson–Sinquefield (IS) model lies in the middle ground between a pure bootstrap and a fully parametric forecast. It resamples historical risk premiums but anchors expectations to the current Treasury yield curve. It also models inflation and real short rates as persistent processes rather than purely random draws. This blend of historical resampling and economic structure is what makes the IS framework robust—and is what has made it a touchstone for long-horizon forecasting for nearly fifty years.

We compare the IS model with the circular block bootstrap (CBB), a fully nonparametric method that preserves short-run time-series dependence without imposing any economic structure. We also apply a global mean adjustment—shifting US return expectations toward the broader international experience—to guard against overextrapolating what may have been an exceptionally favorable century for US markets into the future when making forecasts. The result, using data through 2025 and anchored to the current yield curve, is a forecast of approximately 8.0% nominal and 5.6% real geometric average equity returns over the 2026–50 horizon. Readers interested primarily in the forecasts themselves can skip directly to the final section of this chapter, titled "Forecasts."

The Ibbotson–Sinquefield Simulation Model

The IS methodology was one of the first to show investors not just a single expected return but a full range of plausible long-horizon outcomes, along with the probability of each. To appreciate how the approach works, it helps to understand what made it possible. The release of the Stocks, Bonds, Bills, and Inflation (SBBI) dataset in the mid-1970s was a landmark in financial research. For the first time, investors and researchers had access to a standardized, long-run record of actual US asset-class returns—one stretching back to 1926—that reflected what investors could realistically have earned.

With nearly five decades of consistent data in hand, it became possible to ask a new kind of question: What does history tell us about the range of outcomes a long-term investor might face? The IS framework answers that question by simulating the future using the past. Rather than modeling each asset class in isolation, it draws returns across all asset classes simultaneously, preserving the natural relationships between them—how stocks, bonds, and bills have historically tended to move together or apart. This analysis matters enormously in portfolio construction, where the interplay between assets is at least as important as the return of any single one.

But the IS model does not simply assume the future will resemble the past. Its anchor is the yield curve. Long-term government bond prices embed the market's collective expectations about future interest rates and inflation, and the IS model uses those expectations to set the baseline path for all asset classes. Forward rates extracted from the yield curve determine where returns are expected to go—both in nominal terms and after inflation. History then plays a different but equally important role: It tells the model how returns are likely to vary and fluctuate around that market-implied path. Rather than setting the destination, historical data inform the journey. They capture how volatile markets have been, how asset classes have moved relative to one another, and how far outcomes have tended to stray from expectations. The result is a simulation grounded in where markets stand today, enriched by nearly a century of experience about how they actually behave.

To formalize this intuition, we now describe the structure of the model more precisely. We begin with the role of the yield curve in shaping long-run expectations and then outline how returns are decomposed and simulated within the framework.

Yield Curves and "Consensus" Expectations

A key feature of the IS framework is its explicit use of the government bond yield curve to anchor long-run expectations. At the time of the forecast, the observed Treasury yield curve is treated as a summary of current market information about future interest rates. By extracting one-year forward rates from the yield curve, the model recovers the market's consensus expectations of future short-term nominal interest rates, up to a maturity premium.

Let Y_N denote the yield to maturity on an N-year Treasury bond. The corresponding one-year forward rates $\{F_n\}$ are computed mechanically from the yield curve and interpreted as

$$F_n \approx \mathbb{E}(R_{f,n}) + \hat{\mathbb{E}}(R_H), \tag{20.1}$$

where $\mathbb{E}(R_{f,n})$ is the expected nominal one-year risk-free return n years ahead and $\hat{\mathbb{E}}(R_H)$ is the historical average maturity premium.

Expected nominal short rates are then further decomposed into expected inflation and expected real short rates:

$$\mathbb{E}(R_{f,n}) \approx \mathbb{E}(R_{I,n}) + \mathbb{E}(R_{r,n}). \tag{20.2}$$

Because the yield curve does not specify real rates or inflation, the model uses historical dynamics of real rates to forecast $\mathbb{E}(R_{r,n})$ and adjust the mean of the simulated inflation process so that the resulting nominal short-rate path is consistent with the forward rates implied by the yield curve.

This procedure embodies what Ibbotson and Sinquefield (1976b) described as a "consensus" forecast. The simulated economy evolves according to historically observed dynamics, but its long-run expectations are anchored to information implied by current market prices. Changing the initial yield curve therefore shifts the expected path of inflation, real rates, and expected returns on all asset classes in the simulation.

Decomposition of Returns into Components

Given the expected path of nominal short-term interest rates implied by the yield curve, the IS model constructs asset returns from a small number of economically interpretable components. Rather than modeling each asset class directly, nominal returns are decomposed into inflation, real interest rates, and asset-specific risk premiums.

In the original study (Ibbotson and Sinquefield 1976a), annual nominal total returns were constructed for five US asset categories: common stocks ($R_{m,t}$), long-term government bonds ($R_{g,t}$), long-term corporate bonds ($R_{c,t}$), Treasury bills ($R_{f,t}$), and inflation ($R_{I,t}$). From these series, the following components are defined:

- Inflation, $R_{I,t}$
- The real short rate, $R_{r,t}$
- The equity risk premium, $R_{p,t}$
- The maturity (horizon) premium, $R_{H,t}$
- The default premium, $R_{d,t}$

All these components are defined using exact multiplicative relationships. For example, the equity premium is defined as the excess return on stocks relative to T-bills,

$$R_{p,t} = \frac{1+R_{m,t}}{1+R_{f,t}} - 1, \tag{20.3}$$

and the real short rate is obtained by deflating the bill return by inflation,

$$R_{r,t} = \frac{1+R_{f,t}}{1+R_{I,t}} - 1. \tag{20.4}$$

In the simulation, however, returns are recombined using an additive approximation:

$$R_f \approx R_r + R_I; \quad R_m \approx R_f + R_p; \quad R_g \approx R_f + R_H; \quad R_c \approx R_g + R_d. \tag{20.5}$$

Ibbotson and Sinquefield (1976b) noted that the omitted interaction terms are small at annual frequencies. The additive structure greatly simplifies both the interpretation and implementation of the model while remaining close to the underlying accounting identities.

Time-Series Structure and Persistence

To ensure internal consistency with the yield-curve-implied path of expected short rates, the IS model imposes a simple time-series structure on both inflation and real short rates. The motivation for this addition is straightforward: Inflation and interest rates are known to exhibit persistent behavior over time, with shocks tending to have lasting effects. A purely nonparametric resampling of historical inflation rates would fail to capture this persistence, potentially generating unrealistic sequences of outcomes, such as sharp inflation followed immediately by deep deflation, which are rare in actual macroeconomic data.

Inflation and the real short rate are each modeled as first-order autoregressive, or AR (1), processes:

$$R_{I,t} = \gamma_{I0} + \gamma_{I1} R_{I,\,t-1} + \varepsilon_{I,t}; \quad R_{r,t} = \gamma_{r0} + \gamma_{r1} R_{r,\,t-1} + \varepsilon_{r,t}. \tag{20.6}$$

These AR (1) specifications capture the slow-moving nature of macroeconomic conditions and interest rates, ensuring that simulated paths evolve smoothly around the yield-curve-consistent mean rather than exhibiting implausibly sharp reversals. All coefficients are saved, and the two sets of residuals, $\{\varepsilon_{I,t}\}$ and $\{\varepsilon_{r,t}\}$, are retained for the simulation. These residuals are then resampled during the simulation to generate realistic shocks for inflation and the real short rate, jointly with the fitted coefficients.

By contrast, the equity, maturity, and default premiums are assumed to be serially uncorrelated. In addition, their historical realizations are treated as draws from a stable joint distribution. This approach reflects Ibbotson and Sinquefield's (1976b) view that risk premiums are difficult to forecast through time, even though they exhibit meaningful contemporaneous comovement across asset classes.

Simulation Mechanics and Cross-Asset Dependence

Conditional on the yield-curve-implied path of expected short-term interest rates, the simulation proceeds sequentially, year by year. The initial starting point for inflation and the real short rate is taken from the last observed historical value. Future paths are generated as follows:

- Inflation and the real short rate evolve according to the previously estimated AR (1) processes. Rather than drawing shocks independently, the simulation resamples entire rows of the residual matrix from the AR (1) estimation, so that the inflation shocks, $\{\varepsilon_{I,t}\}$ and $\{\varepsilon_{r,t}\}$, are always drawn from the same historical year, t. This approach preserves their empirical correlation. The maturity premium shock is drawn from the same historical year as well, although this is required only when forecasting returns on longer-duration assets.

- Asset-specific premiums, including equity and default premiums, are likewise drawn together. T-bill returns are constrained to be nonnegative by setting any negative draws to zero and rescaling the remaining values to preserve the historical mean.
- Nominal and real returns for each asset class are then built up from their components using the additive identities described in the previous section.

Repeating this procedure over many simulations produces a large ensemble of plausible future return paths, giving us a distribution of possible return paths over the forecast horizon.

Long-Horizon Outcomes

The simulated return paths are compounded into wealth indices, from which geometric mean returns and percentile bands are computed at each horizon. These outputs form the basis of the familiar fan-shaped “tulip” and “trumpet” plots that summarize the conditional distribution of long-run outcomes implied by the IS model.

Viewed as a whole, the IS model occupies a middle ground between purely historical resampling and fully parametric forecasting. It preserves the empirical joint behavior of asset returns, imposes persistence where economically justified, and disciplines expectations using information from financial markets. Because of the anchoring of expectations to the yield curve, the resulting forecasts are sensitive to prevailing interest rate conditions, allowing the model to adapt to changing economic environments.

Overview of Bootstrap Methods

To understand the IS model’s connection to bootstrap methods, it helps to start with what bootstrapping actually is.

The central idea is simple: Treat the historical record as the best available guide to how returns are generated, and build simulated future paths by repeatedly drawing from that record with replacement. Each simulated path is made up entirely of past observations, reshuffled into a new sequence, with some years possibly appearing more than once and others not at all.

This approach was formalized by Efron (1979) in what is now called the i.i.d. bootstrap—so named because each draw is independent and identically distributed, meaning that past draws have no influence on future ones. By reshuffling historical outcomes, the model constructs a distribution of synthetic future paths without making any assumptions about the underlying data-generating process. It is precisely this resampling mechanism—drawing residuals from the historical record—that the IS model also uses to simulate inflation, interest rate, and risk premium paths. That shared feature is why the IS model has long been associated with bootstrapping.

The difference lies in structure. The i.i.d. bootstrap imposes none: Draws are independent, with no allowance for persistence, trends, or economic relationships. The IS model wraps the same resampling logic inside an economic framework—the AR (1) processes for inflation and real rates and the yield-curve anchor—that gives the simulated paths realistic dynamics.

In portfolio applications, bootstrap methods are typically applied to joint observations across asset classes rather than to individual assets. Drawing an entire year's returns across stocks, bonds, and bills together preserves the historical relationships between them by construction, making the approach well suited to portfolio analysis, where those relationships matter. This empirical grounding, and the fact that no parametric assumptions are required, has made bootstrapping widely used in finance—from studying long-horizon return predictability (Goetzmann 1990; Goetzmann and Jorion 1995) to analyzing tail risk (Daníelsson and de Vries 1997, 2000).

The absence of time-series structure is both the method's strength and its principal limitation. On the one hand, bootstrap forecasts are transparent and rooted directly in the data. On the other, because draws are independent, the simulation cannot reproduce such features as inflation persistence. An i.i.d. bootstrap could plausibly generate a year of sharp inflation followed immediately by deep deflation—a sequence that is rare in practice. This is precisely the gap the IS model closes through its AR (1) structure for inflation and real rates.

A nonparametric alternative exists for closing the same gap: the block bootstrap. Rather than drawing individual yearly observations, block methods resample contiguous segments of the historical record, preserving the short-run time-series dynamics within each block without imposing a parametric model. We focus on two closely related variants: the moving block bootstrap (MBB) and the circular block bootstrap (CBB). Both are well established in the resampling literature (Lahiri 2003), and we use the CBB as a nonparametric benchmark when comparing forecasting approaches in the "Forecasts" section (see also Künsch 1989; Liu and Singh 1992; Politis and Romano 1994).

The Moving Block Bootstrap

The mechanics of this approach are straightforward. Given a time series $\{r_t\}_{t=1}^{T}$ and a block length L, overlapping blocks are constructed by sliding a window across the sample, yielding $T - L + 1$ blocks:

$$\mathcal{B}_1 = (r_1, \ldots, r_L),\ \mathcal{B}_2 = (r_2, \ldots, r_{L+1}),\ \ldots,\ \mathcal{B}_{T-L+1} = (r_{T-L+1}, \ldots, r_T). \tag{20.7}$$

These blocks are then sampled with replacement and concatenated to form a bootstrap path of approximately the original length. Dependence is preserved within each block, although it is broken across blocks. The block length, L, governs the tradeoff: A length of one year results in the i.i.d. bootstrap, whereas a length of five years, for example, captures a full business cycle of dependence. The increased time-series dependence comes at the cost of fewer distinct segments from which to draw.

The Circular Block Bootstrap

The CBB addresses a subtle problem in the moving block bootstrap. The issue in MBB block formation is that observations near the start and the end appear in fewer blocks, introducing a representativeness bias. The CBB treats the series as circular, allowing blocks to begin at any index $t \in \{1, \ldots, T\}$ and wrap around the sample when they reach the end:

$$(r_{T-1}, r_T, r_1, r_2, \ldots). \tag{20.8}$$

This yields T possible starting points rather than $T - L + 1$, giving every observation an equal amount of representation. Thus, the MBB and CBB methods are closely related and, in fact, equivalent asymptotically. In finite samples, however, the CBB typically provides a more uniform representation of the data. For this reason, although the MBB serves as a useful starting point, we use the CBB as the modern comparative methodology for the IS model.

Block bootstrap methods thus occupy a natural middle ground between the i.i.d. bootstrap and the IS model by preserving short-run time-series dependence without imposing parametric structure or yield-curve anchoring. This makes the CBB a useful benchmark for isolating the contribution of economic structure in the IS model, which is how we use it in the "Forecasts" section. We also experimented with hybrid approaches that resample blocks of AR (1) residuals or risk premiums rather than individual draws, but these yielded results nearly identical to the original IS model, confirming that the AR (1) structure already captures the relevant time-series dependence. We therefore focus on the pure CBB and the original IS model in what follows.

Estimation Biases of Historical Data

All forecasting methods that rely on historical data, whether fully nonparametric bootstraps or hybrid frameworks, such as the IS model, share a fundamental limitation: They assume that the past provides a meaningful guide to the future. Simulated outcomes are constructed from the historical observations and therefore inherit both statistical properties and the economic environment of the sample period. The question is not whether historical data are useful—they clearly are—but how representative they are of the future environment we are trying to forecast.

Several distinct sources of bias are relevant when using US historical data to form long-run return expectations. The first concern is survivorship and selection bias. Markets that experienced permanent closure, default, or prolonged stagnation are typically absent from forecasting exercises. By conditioning on the performance of a market that not only survived but thrived, forecasters risk overstating expected returns relative to what a broader set of possible historical outcomes would imply (see, e.g., Brown, Goetzmann, Ibbotson, and Ross 1992; Brown, Goetzmann, and Ross 1995).

A second source of bias is valuation expansion. To see why valuation changes matter for forecasting, it helps to decompose what drives equity returns over time. At its most basic, the total return on an equity investment comes from three sources: the dividend yield, the growth in earnings, and any change in the price investors are willing to pay per dollar of earnings—the P/E. The first two reflect the underlying economics of the business. The third reflects such factors as risk appetite, market participation, investor sentiment, liquidity, and demographics. Over the twentieth century, US equity valuations expanded dramatically. Investors moved from applying relatively low multiples to corporate earnings—reflecting the uncertainties of an earlier era, limited diversification, and less liquid markets—to the substantially higher multiples that prevailed in recent decades. This rerating added a persistent tailwind to realized returns that is visible in the historical record and therefore embedded in any simulation that resamples from it.

The problem is that valuation expansion of this magnitude may not repeat. For instance, a P/E can expand from 10 to 20, generating a powerful boost to returns over the period in which it occurs. But a multiple cannot expand from 20 to 40 with the same ease, and it certainly cannot keep expanding indefinitely. To the extent that future returns revert toward their fundamental drivers—dividend income and real earnings growth—rather than continuing to benefit from further multiple expansion, a bootstrap simulation anchored in the historical record can systematically overstate what investors should expect in the future.

This effect is not small. Empirical decompositions of historical US equity returns suggest that valuation expansion has accounted for a meaningful share of the total realized premium of stocks over bonds. Stripping it out produces a materially lower estimate of the forward-looking equity premium—one more consistent with what the current level of valuations and dividend yields would imply.

More generally, the macroeconomic and institutional stability of the United States, alongside technological progress, favorable demographics, and geopolitical dominance, has created a unique environment over the past century. Forecasts conditioned solely on this experience may embed an optimism that is difficult to disentangle from the data. These biases point to two simple adjustments that can serve as transparent starting points for thinking about how sensitive forecasts are to sample assumptions.

The first is a valuation adjustment that strips out the contribution of P/E expansion from historical returns. One can construct a return series that better reflects what investors might expect if multiples remain stable in the future. In practice, this approach tends to produce forecasts that are modestly lower than the unadjusted US estimates but still reflect the favorable underlying growth experience of the US economy.

The second is a global mean adjustment, which aligns the mean of the US return series with the broader international experience using data from Dimson, Marsh, and Staunton (2026). Rather than replacing US data entirely, the adjustment shifts the mean while preserving the distributional characteristics of the US return series:

$$r_t^{adj} = r_t^{US} - (\mu_{US} - \mu_{WxUS}), \tag{20.9}$$

where μ_{US} is the historical mean of US returns and μ_{WxUS} is the historical mean of world ex-US returns (GDP weighted across countries). This adjustment captures a wider range of economic and political histories, including episodes of war, prolonged stagnation, and institutional disruption that are not present in the US record.

Because the global adjustment produces lower expected returns than the valuation adjustment while remaining very transparent, it serves as the more conservative of the two, and we adopt it as our preferred baseline in the "Forecasts" section. Presenting forecasts relative to this adjustment gives readers a sense of how much of the US historical return premium may reflect conditions that are unlikely to persist. Importantly, neither adjustment should be interpreted as a formal correction; both are better understood as alternative conditioning assumptions about which historical outcomes are most informative regarding the future.

It is worth noting that much more sophisticated adjustments exist. We focus on these two simple adjustments because they are transparent, replicable, and sufficient to illustrate the sensitivity of long-horizon forecasts to sample assumptions.

Forecasts

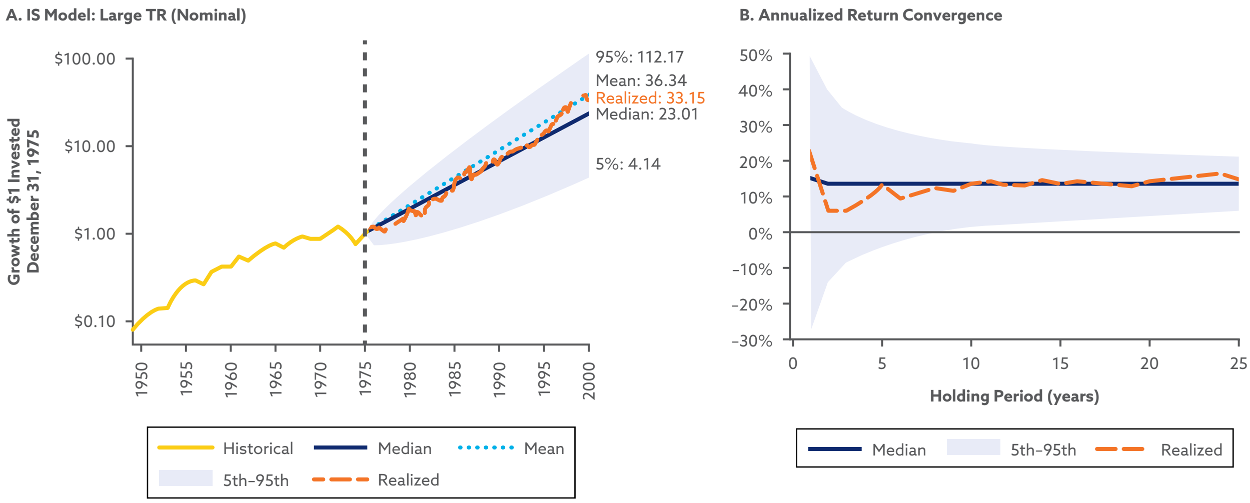
We now present long-horizon forecasts of US equity returns over a 25-year horizon. Each forecast represents a full distribution of potential future outcomes rather than a single point estimate. We summarize these distributions by reporting their mean and highlighting the median and the 5th–95th percentile bands, which together provide a compact description of the potential range of outcomes.

The Original Ibbotson–Sinquefield Forecasts

The original Ibbotson–Sinquefield (1976b) simulations attracted considerable attention both for their novel methodology and their optimistic long-horizon equity return forecasts. Using annual US data from 1926 to 1974 and anchoring expectations to the 1975 yield curve, the authors forecasted a geometric average return of approximately 13% per year for US common stocks over the subsequent 25 years. After adjusting for inflation, this forecast corresponded to an expected real return of roughly 6.3% per year.

Exhibit 20.1 reproduces the original IS forecast for nominal equity returns using the Ibbotson Equity Indices large-cap data. Panel A shows the familiar tulip plot summarizing the distribution of simulated wealth paths, and Panel B presents the corresponding trumpet plot of annualized geometric returns across holding horizons. The realized nominal returns lie close to the center of the simulated distribution for the entire simulation period.

Exhibit 20.1. Replication of the Ibbotson–Sinquefield (1976b) Bootstrap Forecast for US Equities, Anchored to the 1975 Yield Curve, with Realized Returns Overlaid, 1976–2000



Source: Calculated by the authors using Ibbotson Indices data. Underlying data from CRSP and US Bureau of Labor Statistics.

These predictions were later revisited by Ibbotson (2006), who confirmed their exceptional accuracy. Realized returns tracked the forecasted mean nearly identically for the first 20 years before deviating slightly, only to revert back toward the forecast center by the end of the period. Much of the difference between the forecasted and realized results reflects inflation dynamics: The higher-than-expected inflation of the mid-1970s pushed up subsequent nominal forecasts, but as inflation declined sharply over the following decades, realized real equity returns exceeded the model's expectations even as nominal returns tracked the forecasted path closely. The model got the nominal trajectory right for largely the right reasons, even if the underlying real return versus inflation split came out differently than predicted.

The episode offers the right frame for interpreting any long-horizon forecast: The IS methodology is best understood as a statement about the central tendency of a wide distribution—not a guarantee and not immune to history's capacity for surprise. With that backdrop in place, we turn to our own forecast for 2026–2050.

2026–2050 Forecasts

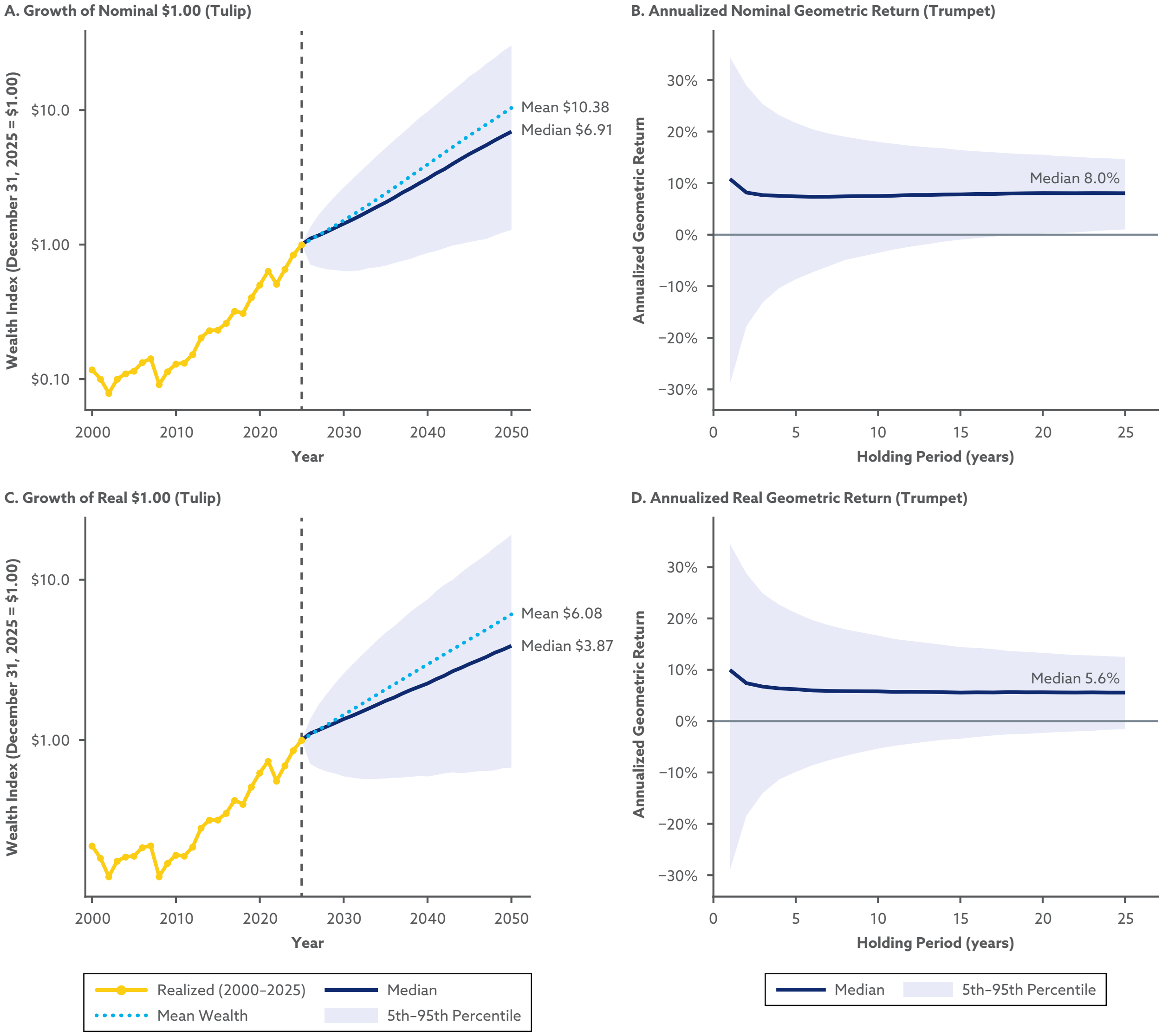
Since the original forecasts proved so accurate over their own 25-year window, we apply the same IS framework to the period 2026–2050 using data from 1926 to 2025, anchoring expectations to the end of 2025 Treasury yield curve. We accompany the IS model forecasts with forecast results generated by the circular block bootstrap (CBB) model as a fully non-parametric comparison, helping assess the role of anchoring expectations to the end of 2025 Treasury yield curve. We run each of the forecast models on two samples: the US record alone and a globally adjusted series. The IS model generates both nominal and real return forecasts, as in the original study, while the CBB is applied only to the real return series due to its indifference about the prevailing interest rate conditions.

On the unadjusted US record, the IS model predicts geometric mean returns of roughly 9.5% nominal and 7.0% real, with the CBB producing a near identical 7.0% real return. Although these figures fall well below the original study—reflecting the change in interest rate environment—many would still call them optimistic, and for a good reason. The very history that these forecasts draw upon—the US experience of the past century—was in many respects an exceptionally fortunate one, and this good fortune is inherited right into our forecasts. We therefore take the globally adjusted IS forecast as our main estimate, one that leans on the broad sweep of global market history rather than the US record alone.

Shifting the sample toward the international experience lowers the forecast by roughly 1.5 percentage points, to approximately 8.0% per year in nominal terms and 5.6% per year in real terms. **Exhibit 20.2** presents the IS forecasts' tulip and trumpet plots using this adjusted US sample. The CBB again tracks the IS model's real figure closely, at 5.5% per year.

The adjustment behind the forecasts retains the distributional shape of the US series, tempering the sample mean toward the broader international experience while leaving its variance intact. The dispersion in the trumpet plots, which tell us the range and relative likelihood of both good and bad outcomes, is therefore still that of the US record, only with the center shifted. This is the same empirical, assumption-free distribution that we highlighted earlier as the strength of non-parametric methods.

Exhibit 20.2. Ibbotson–Sinquefield Bootstrap Forecast for Globally Adjusted US Equities, Anchored to the 2025 Yield Curve, 2026–2050



Sources: Ibbotson Indices; Dimson–Marsh–Staunton global returns database; computations by the authors.

Notes: Panels A and B (top) show nominal return forecasts, with Panel A showing cumulative wealth indices and Panel B compound annual returns; Panels C and D show real return forecasts, with Panel C showing cumulative real wealth indices and Panel D compound annual real returns. The graphs are based on the globally adjusted US return data.

Since the IS and CBB methods produce closely aligned results for the 2026–50 period, the plots between the two are nearly identical. We therefore present the CBB results only as summary statistics in **Exhibit 20.3**. The close agreement between the two methods is partly coincidental and partly structural. At long horizons, short-run time-series dependence—whether captured through the block resampling or an AR (1) process—becomes increasingly irrelevant as returns accumulate, especially because we use annual returns. Thus, the repeated sampling from the same empirical distribution naturally drives estimates toward similar long-run means. In that sense, convergence is expected to an extent.

But the reason that both the IS and CBB forecasts formed using 2025 data are so close to one another is the prevailing macroeconomic environment. US short-term rates and inflation are close to their past-century averages, which means the yield-curve anchoring that distinguishes the IS model from pure resampling happens to add relatively little right now. When the present looks like the historical average, the two approaches rooted to the historical data will mechanically arrive at similar numbers. This is not always the case, however. In 1975, for example, elevated inflation pushed IS nominal forecasts higher, resulting in annual forecasts roughly 4.5 percentage points higher than what the CBB would have predicted. The convergence we see today is therefore as much a statement about current conditions as it is about the methods themselves.

Exhibit 20.3 reports all of the forecasts using both methods (IS and CBB) and both samples (unadjusted US and globally adjusted). The global adjustment lowers the IS forecast by approximately 1.5 percentage points, from 9.5% to 8.0% in nominal terms and from 7.0% to 5.6% in real terms. The CBB real figures also move by the same amount when compared across the two samples. The 8.0% nominal and 5.6% real globally adjusted IS forecast is the one we put forward as our geometric mean forecast for the next 25 years.

We believe these figures are optimistic yet reasonable. A useful way to assess this is to decompose the real forecast into its components. Current real short-term interest rates are

Exhibit 20.3. Forecasted Equity Return Estimates for 2026–2050: Unadjusted and Adjusted Samples

Method	IS	CBB
Unadjusted forecast (2026–2050)		
2026–2050 (Nominal)	9.5%	—
2026–2050 (Real)	7.0%	7.0%
Global returns adjusted forecast (2026–2050)		
2026–2050 (Nominal)	8.0%	—
2026–2050 (Real)	5.6%	5.5%

Sources: Ibbotson Indices; Dimson–Marsh–Staunton global returns database; computations by the authors.

Notes: Reported values are annualized geometric mean return forecasts. “Unadjusted” refers to simulations based on US historical data without international mean adjustment. “Adjusted” refers to simulations incorporating a global mean adjustment.

approximately 0.9% per year.[1] An expected real equity return of 5.6% therefore implies a forecasted equity risk premium of roughly 4.7% over the risk-free rate. This figure fits well in line with the long-run literature. Therefore, we are not forecasting an unusual equity premium for the next quarter century; rather, we are forecasting one that looks very close to what it always has been, and we believe that is the right starting point.

Measuring the Unmeasurable: Indexing and the Reduction of Knightian Uncertainty

There is a deeper reason the adjusted forecast is reasonable, one that connects to the contribution of Ibbotson and Sinquefield. It has to do with how investor attitudes and equity markets have themselves changed over the past century. In equilibrium, expectations about equity returns will influence investor behavior and vice versa. For 1970s investors, the crash of 1929, the volatility of the 1930s, and the sharp market decline after the 1972 oil crisis loomed large in financial market experience. Statistical analysis like that of Ibbotson and Sinquefield had yet to be performed and widely reported, making it difficult for investors to assess the probability of any given outcome. This Knightian uncertainty—the risk of the "unknown unknown"—may explain the high equity risk premium over the fifty years from 1926 to 1975 and motivate a conservative approach to forecasting future returns.

One can debate how much or how little was known about the risk and uncertainty of equity investing between 1926 and 1975. The Great Crash and ensuing Great Depression in the real economy presented investors with a palpable sense that financial life as they knew it might be coming to an end. Radical change was in the air, and although it seemed likely from contemporary accounts that the financial system would survive in some form, a full recovery, a move to new highs, and a superboom later in American history was only one conceivable outcome. Other, less rosy outcomes seemed more likely.

The 1970s were different. Returns were deeply disappointing, there were many corporate bankruptcies and consolidations, and there was civil unrest. But there was little or no talk of dismantling the financial system and starting over. The market would recover, remain low, or go lower, but it would not go away.

Thus, some measure of Knightian uncertainty was resolved in 1926–1975 that didn't need to be resolved in 1976–2026 and still doesn't (looking forward to 2050). This observation supports the idea of assuming a somewhat lower risk premium in the next 25 years than over the last 100 and, in particular, lower than over the first half of the last 100.

Investors today not only have the benefit of a century and more of past market returns but also a statistical toolkit for predicting future return distributions. The empirical project that Ibbotson and Sinquefield began, and that many of the coauthors of this volume have extended, played a catalyzing role in converting market uncertainty into tractable risk metrics. Thanks in large measure to empirical financial research, the long-run record is now far richer than it was when Ibbotson and Sinquefield first ran their simulations. Data collection and modern statistical methods such as bootstrapping and survivorship analysis have accomplished what amounts,

---

[1] As of 31 December 2025.

almost by definition, to the reduction of Knightian uncertainty—that is, quantification of the previously unquantifiable.

Over the subsequent five decades, this resolution of deep uncertainty has made equity markets a normal, expected feature of economic life for households, institutions, and policymakers. The widespread adoption of passive equity investing, using strategies that closely match the capital-weighted portfolio strategies studied in the first SBBI volume, is compelling evidence that the accumulated knowledge about distributions has influenced beliefs and behavior. The first retail index fund also appeared in 1976 but was initially greeted with widespread skepticism; five decades later, index funds have surpassed active managers in US equity assets, more than $11 trillion is benchmarked to the S&P 500, and index and target-date funds are the default vehicles of retirement saving.

Investment trends since 1976 bear directly on our forecast. We cannot simply re-run the 1976 analysis in 2026, because the original study and the movement it launched have changed how investors understand equity risk. Part of the past century's high realized return may well have been compensation for an uncertainty that has since been measured and largely resolved; as that uncertainty was priced in, a somewhat lower required premium may have followed—already embedded in current valuations and in the IS model's yield-curve anchoring. Stocks are no safer in any probabilistic sense; the variance remains substantial. But our forecast of 8.0% nominal and 5.6% real is best understood as the market's own long-horizon consensus, in a world where the distribution of equity returns is no longer an unknown unknown.

Conclusion

Fifty years separate the original Ibbotson–Sinquefield simulations from the forecasts in this chapter, and that distance is itself a measure of how far the discipline of finance has come. In 1976, Ibbotson and Sinquefield used the historical record to simulate the full distribution of outcomes a long-term investor might face, anchored to the prevailing yield curve. We have repeated that exercise with half a century of additional data, modern resampling methods, and the widest international evidence now available—and arrive at a forecast in the same spirit as the original: approximately 8.0% nominal and 5.6% real for US equities over 2026–2050, implying an equity risk premium close to its long-run historical value.

That durability is the truest tribute to the original work. The Stocks, Bonds, Bills, and Inflation project did more than document the past; it helped convert the long-run behavior of diversified portfolios from a matter of judgment into a matter of measurement. The half-century of indexing and institutional adoption that followed is evidence of that achievement and the reason any careful forecaster in 2026 must reckon with it. We offer our numbers in that tradition: not a prediction to be defended to the decimal but a distribution, rather than a point, because the future remains, as it always has been, a range of possibilities rather than a certainty.

Key Takeaways

- Using the Ibbotson–Sinquefield model on the globally adjusted US return history, we forecast 8.0% nominal and 5.6% real annualized geometric returns for US equities over 2026–2050.

- Sampling-based long-horizon return forecasts offer a transparent, data-driven way to form expectations of future equity returns. Their strength is that the path from data to forecast is easy to follow, but this same reliance on data is simultaneously their limitation, as the method implicitly assumes that the past is a reasonable guide to the future. The forecasts thus are only as representative as the sample they draw from.
- The Ibbotson–Sinquefield model highlights how incorporating a modest amount of economic structure, such as yield-curve anchoring and persistent inflation dynamics, can produce more economically grounded forecasts. Such forecasts come at the cost of being more sensitive to prevailing market conditions.
- Fully nonparametric block bootstrap methods offer a complementary approach that can also preserve time-series dependence without imposing parametric assumptions.
- The forecast based on US historical data was adjusted by lowering the expected equity risk premium to the international historical equity risk premium mean. The reason for this adjustment is that the US was the best performing equity market, with the US market benefiting during the last 100 years from survivorship, valuation, and institutional circumstances that are not likely to repeat.

References

Brown, Stephen J., William N. Goetzmann, Roger G. Ibbotson, and Stephen A. Ross. 1992. "Survivorship Bias in Performance Studies." *Review of Financial Studies* 5 (4): 553–80.

Brown, Stephen J., William N. Goetzmann, and Stephen A. Ross. 1995. "Survival." *Journal of Finance* 50 (3): 853–73.

Daníelsson, Jón, and Casper G. de Vries. 1997. "Tail Index and Quantile Estimation with Very High Frequency Data." *Journal of Empirical Finance* 4 (2–3): 241–57. doi:10.1016/S0927-5398(97)00008-X.

———. 2000. "Value-at-Risk and Extreme Returns." *Annales d'Economie et de Statistique* 60: 239–70. doi:10.2307/20076262.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2026. *Global Investment Returns Yearbook 2026*. UBS Global Wealth Management. www.ubs.com/global/en/wealthmanagement/insights/2026/global-investment-returns-yearbook.html.

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics* 7 (1): 1–26. www.jstor.org/stable/2958830.

Goetzmann, William N. 1990. "Bootstrapping and Simulation Tests of Long-Term Patterns in Stock Market Behavior." PhD dissertation, Yale University.

Goetzmann, William N., and Philippe Jorion. 1995. "A Longer Look at Dividend Yields." *Journal of Business* 68 (4): 483–508.

Ibbotson, Roger. G. 2006. "Predictions of the Past and Forecasts for the Future: 1976–2025." In *The Equity Risk Premium: Essays and Explorations*, edited by William N. Goetzmann and Roger G. Ibbotson, 266–69. Oxford University Press. doi:10.1093/oso/9780195148145.003.0013.

Ibbotson, Roger G., and Rex A. Sinquefield. 1976a. "Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926–1974)." *Journal of Business* 49 (1): 11–47. doi:10.1086/295803.

———. 1976b. "Stocks, Bonds, Bills, and Inflation: Simulations of the Future (1976–2000)." *Journal of Business* 49 (3): 313–38. doi:10.1086/295854.

Künsch, Hans R. 1989. "The Jackknife and the Bootstrap for General Stationary Observations." *Annals of Statistics* 17 (3): 1217–41. www.jstor.org/stable/2241719.

Lahiri, Soumendra Nath. 2003. *Resampling Methods for Dependent Data*. New York: Springer.

Liu, Regina Y., and Kesar Singh. 1992. "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence." In *Exploring the Limits of Bootstrap*, edited by R. Lepage and L. Billard, 225–248. New York: Wiley.

Politis, Dimitris N., and Joseph P. Romano. 1994. "The Stationary Bootstrap." *Journal of the American Statistical Association* 89 (428): 1303–13. doi:10.1080/01621459.1994.10476870.

CHAPTER 21. CONCLUSION

Roger G. Ibbotson, PhD
Professor in the Practice Emeritus of Finance, Yale School of Management
Chair, Zebra Capital Management, LLC

In 1976, Rex Sinquefield and I published “Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926–1974)” in the University of Chicago’s *Journal of Business* (Ibbotson and Sinquefield 1976a). This year marks the 50th anniversary of that publication and the 100th anniversary of the 1926 beginning of the Stocks, Bonds, Bills, and Inflation (SBBI) dataset. This year, my coauthors and I introduce a new dataset that we hope will carry the methods and the thinking behind SBBI forward into the future.

Ibbotson Associates updated SBBI continually until 2006, when the firm was sold to Morningstar, Inc. SBBI was updated by Morningstar through 2024, until it was discontinued in early 2025.

The Ibbotson Equity and Bond Indices in this book replace and expand the original SBBI indices. Morningstar announced that it was acquiring CRSP (Center for Research in Security Prices) in September 2025, and the deal closed in February 2026. The indices presented in this book use CRSP (with permission from CRSP) as the underlying data source. Until its acquisition, CRSP was created and affiliated with the University of Chicago’s Booth School of Business. I previously served as executive director of CRSP from 1979–1984 while I was a faculty member at Chicago Booth.

In Chapter 2, Tom Idzorek, Larry Siegel, and I describe the summary results of the Ibbotson Equity and Bond Indices over the last 100 years. From 1926 to 2025, the compound total annual return on large-cap stocks was 10.1%. Small caps had a 10.9% return. Long-term US Treasury bonds returned 4.9%, while 30-day Treasury bills returned 3.3%. Inflation was 2.9% over the period.

The partial title of this book is *Exponential Wealth*, reflecting the fact that when total returns are compounded over a long period, with income reinvested and before any taxes or expenses are removed, tremendous wealth can be created. Over the last 100 years, $1 invested at 10.1% in large-cap stocks grew to $14,751. Even after inflation, which caused prices to rise 18-fold, the real value of the original $1 investment is $820. Such exponential wealth does not emerge from the lower returns of bonds and bills: $1 grew in real (inflation-adjusted) terms only 6.5 times for long-term US Treasury bonds and 1.39 times for US Treasury bills.

In general, the higher the risk, the higher the return. Over this 100-year period, there was a positive equity risk premium, horizon (interest rate risk) premium, and small-cap premium, as well as a positive real riskless interest rate. In Chapter 9, Larry Siegel and Carla Nunes study the equity risk premium in detail with various estimation methods. The math and index construction techniques are presented by Jim Harrington and me in Chapter 3.

The US Equity Indices

Otto Manninen, Will Goetzmann, and I computed the Ibbotson Equity Indices from the CRSP data, and we present the results in Chapter 4. The Ibbotson Indices are a set of capitalization-weighted indices that include the top 25 stocks (Mega Cap), Large Cap, Mid Cap, Small Cap, and Micro Cap, along with composite indices that combine these elemental pieces. During the first half of the study period (1926–1975), the cuts are based on decile counts, as the number of stocks in the underlying dataset grow dramatically over time. At the start of the sample in 1926, there were only a few more than 500 eligible stocks in the CRSP database. Thus, Deciles 1 and 2 (large caps) contained, between them, only about 100 stocks. By 1975, after AMEX and Nasdaq stocks had been added to the CRSP database, there were more than 4,600 stocks, with the top two deciles consisting of more than 920 stocks. The large-cap index thus represents the dominant companies of the time, with the same relative positions in terms of importance to the economy.

For the later period, 1976 to present, the index cuts are based on fixed counts. The CRSP datasets continued to grow, reaching a maximum of more than 7,500 stocks in the 1990s. Although the number of stocks has fallen in this century, there are still almost 4,000 stocks today, a sufficient number to provide fixed-count cut-offs. The Large Cap index, for example, consists of the top 300 stocks by capitalization. This number is held constant despite changes in the total count of US stocks being traded on the various exchanges. This is the way that many of today's popular indices, such as the S&P 500 and the Russell 2000, are constructed.

During the full 100-year period, compound annual total returns were 10.07% for large caps, 11.05% for midcaps, 10.95% for small caps, and 11.06% for microcaps. The small-cap premium over large caps is positive, but it is not consistent across all size categories and amounts to only about 1% per year.

In Chapter 5, Will Goetzmann and Otto Manninen analyze the US indices, noting that they are positively skewed with kurtosis, meaning that a few winning stocks account for much of the equity risk premium. These authors, along with coauthor James Tyler, follow up, in Chapter 6, with a study of bubbles, booms, and crashes from 1792 to the present. The authors find that crashes have been rare, mostly unpredictable, and usually followed by recovery rather than a further crash, a finding that enriches our theme of exponential wealth being created by long-term investing in equities.

The US Bond Indices

The Coleman Fisher Ibbotson (CFI) Bond Indices consist of estimated forward interest rates, yield curves, and total returns on US Treasury issues over the period 1926–2025. In Chapters 7 and 8, Tom Coleman and I discuss these yield curves, returns, and methodology. By estimating forward rates, the CFI data can be used to construct any type of yield curve (e.g., par bond, annuity, zero) and price any set of default-free cash flows over the last 100 years. Furthermore, synthetic bond returns of any maturity or duration can be created.

The 100-year history of US Treasury yields can be summarized as follows: Yields fell in the 1930s and then rose, peaking in 1981 and falling thereafter. A graph of the yields looks like a simple, predictable pyramid, with both short- and long-term yields peaking at more than 15% in 1981, although the month-to-month yield changes are nearly random. Because of the rise in yields before 1981, investors experienced capital losses, and because of the later decline in yields, they had capital gains. For the whole period, however, capital gains/losses were near zero, as yields coincidentally started and ended the 100-year period near 4%.

The overall average yields were 3.25% for short-term bills (one month) and 4.86% for long bonds (20 years), providing a historical horizon premium of 1.61%. Because of the rise in yields, the 1925–80 average total return of the long bond was only 2.69%. Because of the later decline in yields, the 1981 to present long bond return averaged 7.62%. Over the entire 100-year period, most of the return came from the yield, whereas most of the volatility came from the change in the yields.

This book features numerous coauthors who have created datasets over the centuries for US and global markets.

Global Markets and Older Data

Ed McQuarrie, in Chapter 10, has backdated the US stock and bond data to 1793, shortly after the Buttonwood Agreement of 17 May 1792, which created the New York Stock Exchange (NYSE). He expands the earlier Goetzmann, Ibbotson, and Peng (2001) work by including stocks beyond the NYSE, as well as additional dividend and capitalization data. His data can potentially be linked to the Ibbotson Indices, starting in 1926. During this pre-1926 period, the United States generally had deflation between wars. US bonds were not riskless but outperformed US equities before the Civil War and almost matched equities for the overall period up to 1926.

In Chapter 11, Elroy Dimson, Paul Marsh, and Mike Staunton provide extensive real total returns for 21 countries, starting in 1900. They show that stocks outperformed bonds and bonds outperformed bills in almost every country. However, the United States, which currently makes up more than half of the global market, is not very representative of the total, because it is the best-performing market over the 126-year period. Later, we adjust our forecasts to reflect the generally lower global returns outside the United States.

The historical London market had a similar experience. In Chapter 12, Will Goetzmann, Fernando Reyes De La Luz, and Geert Rouwenhorst show that during the period from 1870 to 1929, the equity premium over long-term bonds averaged 4.5% per year, with realized returns declining monotonically with the seniority of the asset class in a typical firm's capital structure.

Bryan Taylor also examines stock and bond data in earlier centuries in Chapter 16, summarizing the results of the vast Finaeon dataset. He identifies four financial revolutions, starting with the first Venetian loans in 1171 and then the founding of four East Indian trading companies beginning in the 1600s. He moves on to global stock and bond trading in the 1800s and finally covers the more recent explosion in global issuance and investing.

This book also examines other markets. Geert Rouwenhorst, Rajkumar Janardanan, and Xiao Qiao created a suite of commodity futures. They show in Chapter 13 that the performance of futures returns was distinct from spot returns and that futures were competitive with US stock markets over the period 1871–2025, providing diversification benefits.

Tadaaki Komatsubara, in Chapter 14, measures the remarkable rise in Japanese markets from 1952 to the late 1980s, which ended in a bubble, followed by a two-decade adjustment and a recent recovery. In Chapter 15, Peng Chen documents the Chinese market over 2005–2025, with returns that closely align with longer-term studies of US markets. These results affirm that exponential wealth can be built in a variety of markets around the world, albeit at different rates and times.

David Chambers, Elroy Dimson, Antti Ilmanen, and Paul Rintamäki evaluate and critique the whole index literature in Chapter 17. The authors highlight the challenges: easy-data biases, macro-consistency, replicability, and total return estimation. The authors of Chapter 17 mostly meet these challenges, but of course, index construction is an ongoing learning process as we find out more and more about both markets and history.

The Future

In Chapter 18, "Popularity and Premiums," Tom Idzorek, Paul Kaplan, and I summarize some of our earlier work on how assets are priced—that is, how expected returns are determined. Assets with disliked characteristics (e.g., the assets are risky, lack liquidity, are boring, have a long duration, or are from polluters) are unpopular, with low current prices and relatively high expected returns. The opposite is true for popular assets, with higher prices and lower expected returns.

Paul Kaplan models the future distributions of the various asset class returns in Chapter 19 using statistical tools. He estimates the first four moments of the distribution of stocks, bonds, and bills—that is, the mean, standard deviation, asymmetry (skewness), and kurtosis. He shows that although lognormal distributions work well for diversified portfolios, asymmetry and kurtosis are required to build satisfactory models of specific asset classes, especially the riskier ones.

The book ends with long-term forecasts of the stock and bond markets. In Chapter 20, Will Goetzmann, Otto Manninen, and I use the original Ibbotson and Sinquefield (1976b) estimation methodology to forecast the stock market over 2026–2050 using bootstrapping techniques that incorporate the inflation and real interest rate expectations embedded in the current year-end 2025 yield curve. We then add the full distribution of equity risk premiums derived from the 100-year historical samples.

We do make an adjustment, however. We recognize that, as Dimson, Marsh, and Staunton noted in Chapter 11, the United States was an unusually successful country over the last century. We therefore subtract the return of the US stock market in excess of the GDP-weighted non-US market (about 1.5% per year). Our forecast for the US stock market over the period 2026–2050 is thus 8% nominal and 5.6% real (after inflation).

I hope that future researchers will continue to work on stock and bond total return indices for as long as capital markets continue to exist. The capital markets are vitally important because they enable wide and deep participation in the exponential wealth-building engine documented in this book.

References

Goetzmann, William N., Roger G. Ibbotson, and Liang Peng. 2001. "A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability." *Journal of Financial Markets* 4 (1): 1–32. doi:10.1016/S1386-4181(00)00013-6.

Ibbotson, Roger G., and Rex A. Sinquefield. 1976a. "Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926–1974)." *Journal of Business* 49 (1): 11–47. doi:10.1086/295803.

———. 1976b. "Stocks, Bonds, Bills, and Inflation: Simulations of the Future (1976–2000)." *Journal of Business* 49 (3): 313–38. doi:10.1086/295854.