SECTION 2

The Past 100 Years in US Markets

CHAPTER 4. A CENTURY OF US EQUITY INDICES

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

Otto S. Manninen
Research Associate, Yale School of Management

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

Introduction

Historical market data have been the statistical foundation for evaluating the risk and return of investing since the emergence of modern portfolio theory. Long-run return series provide the empirical basis for asset allocation, financial planning, capital budgeting, and portfolio construction. For equity investors in particular, understanding how markets have behaved across different economic regimes is essential for forming expectations about the future.

In this chapter, we introduce a new family of US equity indices constructed from security-level data provided by the Center for Research in Security Prices (CRSP). These indices span the full 1926–2025 period and are designed to represent capitalization-weighted portfolios segmented by firm size. The goal is to provide a transparent, consistent, and economically meaningful representation of the US equity market over a full century.

The century from 1926 to 2025 includes extraordinary periods. It spans the Great Depression, World War II, the postwar expansion, the inflationary shocks of the 1970s, the technology boom and bust, the global financial crisis, and the COVID-19 pandemic, while also capturing the beginning of the current AI-driven technology boom. During the course of the century, the structure of the economy evolved dramatically, with regimes ranging from industrial manufacturing dominance to a technology- and services-oriented market. Yet despite these changes and challenges, US equities delivered extraordinary long-run returns, even when compared with the global markets (Dimson, Marsh, and Staunton 2002).

Over the full 100-year sample, the broad composite market portfolio produced a compound average total return of approximately 10.3% per year. Large-cap stocks delivered an average compound return slightly above 10%, while midcap, small-cap, and microcap portfolios generated even higher long-run compound returns, around 11% annually. These differences may appear modest on an annual basis, but over multidecade horizons, they compound into large differences in terminal wealth.

To illustrate, without taxes or other costs, $1 invested in 1926 in the broad market portfolio would have grown to more than $17,000 by the end of 2025, while the smaller size segments compound to roughly double that. Even after accounting for inflation, the real increase in purchasing power remains substantial. The century-long record demonstrates both the resilience and volatility of public equity markets. Severe drawdowns occurred, but recoveries followed, and long-term investors were rewarded for remaining invested.

At the same time, the distribution of returns differs meaningfully across size segments. Smaller-cap portfolios historically delivered higher average returns but also greater volatility. Large-cap portfolios exhibited somewhat lower return dispersion and, in some subperiods, higher risk-adjusted performance. Understanding these differences is central to both strategic asset allocation and portfolio diversification.

The indices presented in this chapter are constructed to facilitate such analysis. They are purely market-cap weighted and use stable, rank-based size definitions. Unlike many modern commercial benchmarks, they do not apply free-float adjustments. This design choice preserves methodological consistency over the full 1926–2025 period and ensures comparability across structural shifts in the market.

In this chapter, we first describe the construction of the Ibbotson Equity Indices, explaining key methodological choices, including capitalization weighting and the distinction between full-cap and float-adjusted approaches. We then examine the century-long behavior of returns across size segments and discuss why long-run data remain essential for estimating expected returns and risk.

Construction of the Ibbotson Equity Indices

The Ibbotson Equity Indices are designed to provide a transparent and consistent representation of the US equity market across the full last century. The construction methodology has been selected to emphasize transparency and replicability, with a key focus on long-run consistency. All portfolios are fully based on market capitalization of the underlying firms.

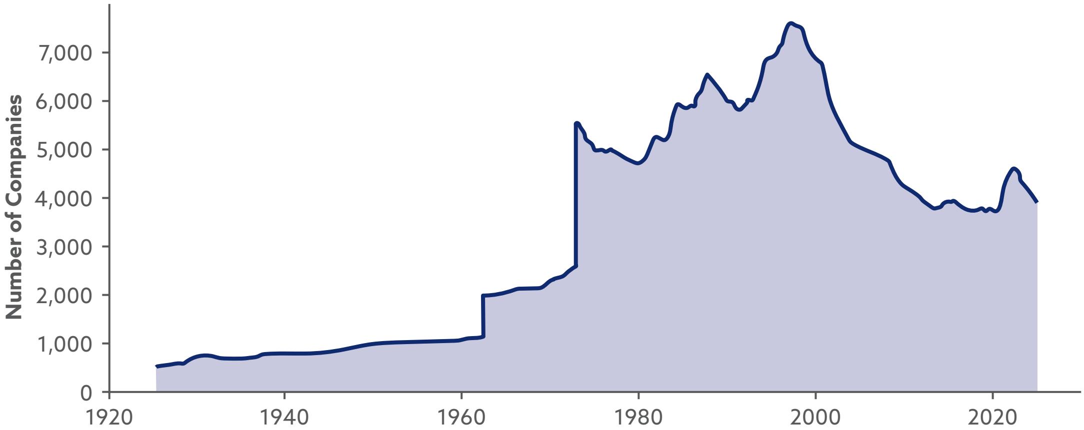
The objective is not to replicate any specific commercial index. Rather, the Equity Indices are constructed to provide economically meaningful size-segmented portfolios that can be used for long-run analysis, benchmarking, and research. The focus is on preserving stable definitions across time so that comparisons across decades reflect economic change rather than methodological revisions.

Underlying Universe

All the Equity Indices are constructed from the CRSP Monthly Stock File, which provides security-level return and market capitalization data beginning in 1926. This dataset offers a comprehensive view of the development of US equity markets and accounts for several biases that can arise in historical return series.

In particular, CRSP data address survivorship bias by tracking firms after delisting and incorporating the last return after delisting, where available. This process ensures that returns reflect the experience of investors holding securities through such corporate events as mergers, liquidations, and bankruptcies. More broadly, CRSP provides comprehensive historical coverage,

Exhibit 4.1. Number of Companies Listed on the NYSE, Nasdaq, and American Exchanges, 1926–2025



Source: Ibbotson Equity Indices, underlying data from CRSP.

Note: The sudden changes in 1962 and 1973 are due to the addition of the American Stock Exchange and Nasdaq, respectively, to the CRSP database and do not reflect changes in the number of US publicly traded firms in existence.

accurate security identifiers, detailed corporate action adjustments, and well-documented methodology. For these reasons and because of the overall depth and quality of the dataset, the CRSP universe is widely regarded as the gold standard in academic research, providing a consistent representation of the US public equity market.

The universe used for index construction includes common stocks listed on the NYSE, Nasdaq, and NYSE American (formerly the American Stock Exchange, widely known as AMEX). We constrain the securities included to the common stocks listed, excluding any exchange-traded funds, REITs, closed-end funds, and American depositary receipts from the final universe used in index construction.

The number of publicly traded firms has changed significantly over time, and these changes have been well documented (Doidge, Karolyi, and Stulz 2017). As shown in **Exhibit 4.1**, the number of stocks in the US equity universe expanded steadily through much of the 20th century, peaking in the late 1990s and declining in the subsequent decades.

These dynamics have been a major influence in the index construction methodology, motivating adjustments to portfolio breakpoints across subperiods.

Classifying Stocks into Size Categories

The Ibbotson Equity Indices use two approaches to size segmentation, each reflecting the evolving structure in the US equity market. Before 1975, we rely on the preexisting CRSP decile portfolios for all indices except the Ibbotson 25, which is constructed directly from firm-level rankings to capture the 25 largest companies at each point through the entire century.

During the early decades of the sample, the equity market was rapidly evolving and the number of stocks listed increased quickly, including the addition of Nasdaq-listed securities in 1972. Using the CRSP decile definitions during this period provides a stable and consistent classification framework while the market structure matured.

After 1975, the Equity Indices are partitioned using fixed numerical market-cap rank thresholds. Using fixed rank thresholds offers key advantages. Most notably, it provides intuitive and transparent definitions, such as "top 500 firms," that are already commonly used among some of the largest commercial indices around the world.

The full equity index definitions are listed in **Exhibit 4.2**.

These definitions ensure that the indices remain economically comparable across time while accommodating structural changes in market breadth. Before 1975, CRSP deciles provide stable classification during a period of rapid change in the US public markets. After 1975, fixed rank thresholds offer intuitive and transparent segmentation.

Exhibit 4.2. Definitions of the Ibbotson Equity Indices

A. Base Indices		
Index Name	**1926–1975 Definition**	**1976–2025 Definition**
Ibbotson 25 (Mega Cap)	Ranks 1–25	Ranks 1–25
Ibbotson Large Cap	Deciles 1–2	Ranks 1–300
Ibbotson Mid Cap	Deciles 3–5	Ranks 300–1,000
Ibbotson Small Cap	Deciles 6–8	Ranks 1,000–2,000
Ibbotson Micro Cap	Deciles 9–10	Ranks 2,000+
B. Composite Indices		
Index Name	**1926–1975 Definition**	**1976–2025 Definition**
Ibbotson Large Composite	Deciles 1–5	Ranks 1–1,000
Ibbotson Small Composite	Deciles 6–10	Ranks 1,000+
Ibbotson SMID Composite	Deciles 3–10	Ranks 300+
Ibbotson All-Cap Composite	CRSP Market Portfolio	Entire universe

Capitalization Weighting

Within each size segment, securities are weighted by total market capitalization. Portfolio constituents are picked quarterly based on the previous definitions. Because all the portfolios are purely market-cap weighted, the individual weights need no rebalancing in response to ordinary price changes.[1] As a result, the portfolios remain aligned with market valuations and any major changes to portfolio weights occur only during the predetermined quarterly rebalancing dates.

When compared with many commercial indices, a major difference in the Ibbotson Equity Index portfolios is that they are constructed using full market-cap weighting instead of float-adjusted weighting. Float-adjusted weighting is very common among many modern commercial benchmarks and indices. In this weighting scheme, market value is calculated using only the shares available for public trading and excludes insider or otherwise restricted holdings. Float weighting can improve investability for index-tracking products by more closely matching tradable supply.

Consistent float data are not available for the full 1926–2025 period, however. Applying float adjustments only in recent decades would introduce a structural break in the methodology. Because the priority of the Equity Indices is long-run consistency, we have opted for full market-cap weighting over the entire period. This way, the indices also reflect the full economic footprint of publicly traded firms, rather than only the investable float-adjusted subset.

In practice, for many firms, the difference between total market capitalization and float-adjusted capitalization is modest, often in the single-digit percentage range. In extreme cases where insider ownership is substantial, the difference can be larger, in the 10%–20% range (large differences are more common for non-US securities). Over long horizons, however, these differences are generally small relative to overall market movements.

The choice to use full market-cap weighting can thus cause deviations from modern float-adjusted commercial benchmarks. A more appropriate set of benchmarks for the new Equity Index series can be found in other research indices, which typically do not use float adjustments. Over long horizons, differences between full-cap and float-adjusted weighting tend to be modest, with long-run mean returns and standard deviations nearly identical for the two methods.

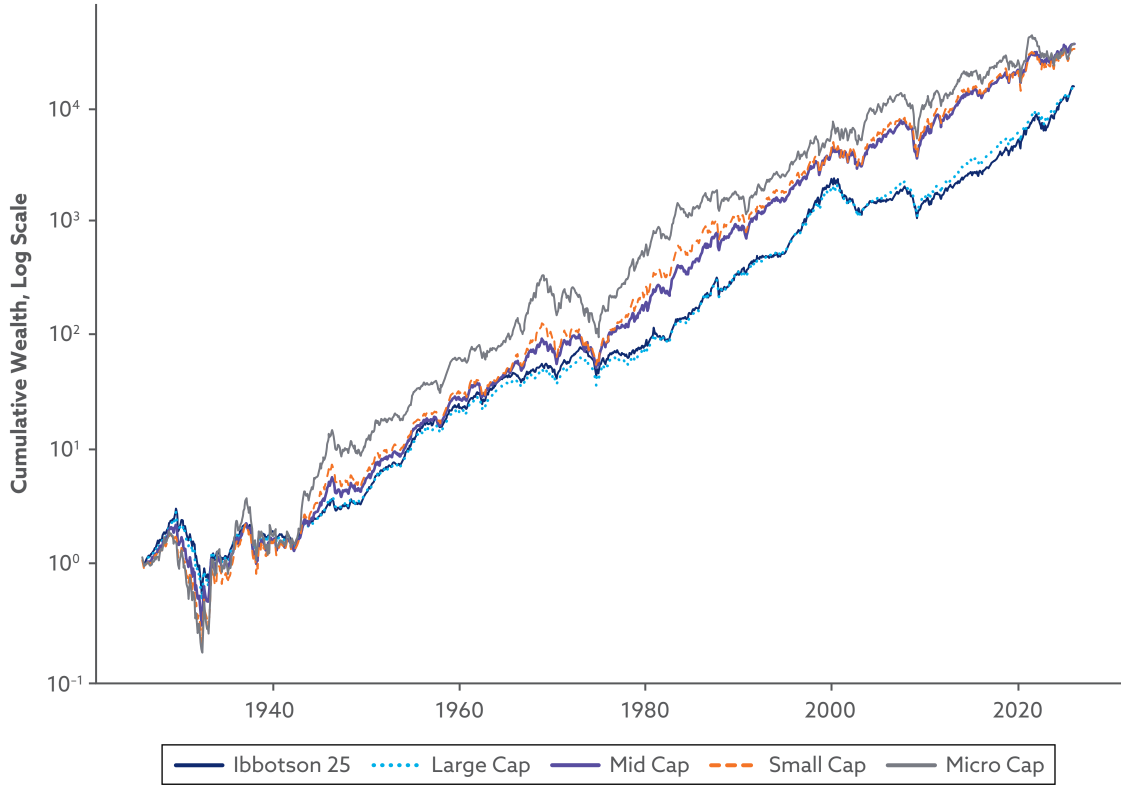
Results: A Century at a Glance

In this section, we take a wide-angle look at the past 100 years of US equity returns using the new Ibbotson Equity Indices. Our goal is to show the long-run level and variability of returns and provide a common frame of reference for investors and others whose primary focus is compounded portfolio growth and volatility. Thus, we focus on market-wide patterns and sustained trends, leaving in-depth statistical analysis for later.

We begin with the growth of $1 invested at the start of the sample period across our baseline indices of stocks sorted by market cap into the Ibbotson 25, Large Cap, Mid Cap, Small Cap, and Micro Cap categories. **Exhibit 4.3** plots each wealth index, using a logarithmic scale to highlight relative performance and periods of divergence and convergence between different size portfolios.

[1] Only such corporate actions as mergers, divestitures, and buybacks require changes in the weights between quarterly rebalancings.

Exhibit 4.3. Cumulative Wealth Indices of Size Categories of US Stocks, 1926–2025



Source: Ibbotson Equity Indices, underlying data from CRSP.

One of the useful features of the log scale is that in a graph of the growth of wealth over time, the slope indicates the compound rate of return identically no matter where you are in the time period studied (in other words, if a given slope means a 20% return in a year in the 1920s, it means the same thing in a year in the 2020s). An arithmetic scale, in contrast, gives the appearance of all the growth taking place at the end, although that is not in fact what happened.

The results over the past century vividly illustrate the power of compounding. Small, persistent return differences between size portfolios accumulate into large gaps over decades, made visible by the ending values of the cumulative wealth index of each portfolio. The upward log path of the wealth index is nearly linear, implying a roughly constant compound rate over long stretches of time, punctuated by a few dramatic events: the crash of 1929–1932 and the subsequent depression years of the 1930s, market declines in the recession of 1973–1974, the dot-com collapse of 2000–2002, and the global financial crisis of 2007–2009. The crash of 1987—the largest one-day drop in US stock market history—is a relatively tiny blip, as is the spring 2020 COVID-19 market meltdown. Exhibit 4.3 demonstrates that these sharp drawdowns were followed by multiyear recoveries, even though full recovery and a move to new highs sometimes arrived slowly, taking 16 years in the case of the 1929–1932 crash and the Great Depression.[2]

Summary Statistics

In **Exhibit 4.4**, we report summary statistics for the four base indices: Large Cap, Mid Cap, Small Cap, and Micro Cap. For each portfolio, we include the compounded (geometric) and arithmetic annual mean return, the annualized standard deviation of monthly returns, and, alongside, a histogram of the annual return distribution of each index. We display results for the full century, the first half (1926–1975), and the second half (1976–2025) separately. The data illustrate how the risk and return of the various portfolios have changed between different periods and how, over a sufficiently long horizon, US equities have delivered strong real returns across all size categories.

Across different portfolio sizes, smaller companies have on average outperformed their large-cap counterparts—a well-known fact about historical equity performance first documented by Rolf Banz (1981) and Marc Reinganum (1981) and later formalized by Fama and French (1993). What is perhaps surprising, however, is that the midcap portfolio matches or exceeds the small-cap portfolio on a compounded annual basis over all three sample periods. This result is largely a consequence of volatility drag: The higher standard deviations (higher risk) of the smaller portfolios erode their compound returns relative to their arithmetic means, with the gap widening substantially as we move down the size spectrum.

Another factor at play is the definition of small caps. Early treatments of the size premium typically used the small-cap definitions from such datasets as those of Ibbotson and Sinquefield (1976), who defined the small-cap portfolio as Deciles 9–10, corresponding to what we classify here as microcap. When this definitional shift is accounted for, the historical small-cap premium looks considerably more concentrated in the extreme tail of the size distribution, consistent with Banz's (1981) original findings that the effect was most pronounced among the very smallest firms. Notably, in the post-1976 period, the microcap portfolio achieved the lowest compound return of any base portfolio despite being the most volatile—a pattern that challenges the simple narrative of smaller stocks producing a higher long-run reward.

More broadly, both the return premium across size segments and the volatility of individual portfolios in the first and second halves of the sample look meaningfully different, suggesting that the structural features of the US equity market, rather than size alone, play an important role in shaping long-run outcomes. We examine these structural shifts in the sections that follow.

Exhibit 4.5 reports summary statistics for the composite indices, which aggregate the size-sorted segments into broader portfolio groupings. The composites largely confirm the patterns observed among the base indices, sitting between their component building blocks in both return and risk. The Small Cap Composite, which combines small-cap and microcap stocks, and the SMID Composite, which spans midcap, small-cap, and microcap stocks, show how blending size segments smooths out some of the extreme volatility associated with the smallest stocks individually, without meaningfully altering the return profile relative to the base indices. Across all composites, the arithmetic–geometric spread widens with volatility, showcasing the role of volatility drag discussed earlier.

[2] In real total return terms, the 1937 high briefly surpassed the 1929 high, but the market then declined sharply again and did not fully recover until after World War II.

Exhibit 4.4. Summary Statistics and Distributions of Annualized Returns on Principal Asset Classes

	Geometric Mean (%)	Arithmetic Mean (%)	Std. Dev. (%)
A. Full Sample (1926–2025)			
Large Cap	10.07	11.85	19.04
Mid Cap	11.05	13.64	23.68
Small Cap	10.95	14.44	27.69
Micro Cap	11.06	16.52	37.28
Inflation	2.94	3.01	3.92
B. First Half (1926–1975)			
Large Cap	8.24	10.41	21.18
Mid Cap	9.29	13.01	28.83
Small Cap	9.31	14.48	34.09
Micro Cap	10.70	19.11	47.35
Inflation	2.29	2.40	4.79
C. Second Half (1976–2025)			
Large Cap	11.95	13.29	16.72
Mid Cap	12.84	14.28	17.34
Small Cap	12.61	14.40	19.66
Micro Cap	11.42	13.93	23.49
Inflation	3.59	3.63	2.70

Sources: Equity data from the Ibbotson Equity Indices, underlying data from CRSP; inflation data from the US Bureau of Labor Statistics.

Market Concentration and Diversification

One of the most practically relevant insights from a century of size-segmented return data is that “investing in the broad market” has meant something different at different points in history. Today, a passive investor holding a simple cap-weighted US equity index is, by construction, making a far more concentrated bet than their counterpart in 1980—not because they chose to but because the market itself has become much more concentrated.

Because the Ibbotson Equity Indices are capitalization weighted, changes in the relative market value of the largest firms translate directly into changes in index weights because security

Exhibit 4.5. Summary Statistics of Annualized Returns for Composite Indices, 1926–2025

	Geometric Mean (%)	Arithmetic Mean (%)	Std. Dev. (%)
Full Sample (1926–2025)			
All Cap Composite	10.27	12.16	19.59
Large Cap Composite	10.25	12.09	19.30
SMID Composite	10.96	13.84	24.91
Small Cap Composite	10.84	14.66	29.09

Source: Ibbotson Equity Indices, underlying data from CRSP.

weights evolve continuously with market prices without requiring discretionary rebalancing. Changes in index constituents (firms crossing size thresholds) occur only at predetermined quarterly rebalancing dates. Whenever a handful of companies grow to dominate a large share of total market capitalization, a passive investor's portfolio tilts toward them automatically and unavoidably.

Exhibit 4.6 shows how the market-capitalization shares of the five base index segments—Ibbotson 25 (Mega Cap), Large Cap (includes Ibbotson 25), Mid Cap, Small Cap, and Micro Cap—have evolved over 1926–2025. The share of the top 25 firms has fluctuated between roughly 20% and 50% of total US market capitalization over the past century. Concentration was elevated in the early part of the past century, declined steadily from the mid-1970s through the early 2000s as the number of listed firms broadened dramatically, and has risen sharply since then. Today, the top 25 firms account for approximately 50% of total market capitalization—a level not seen since the 1930s.

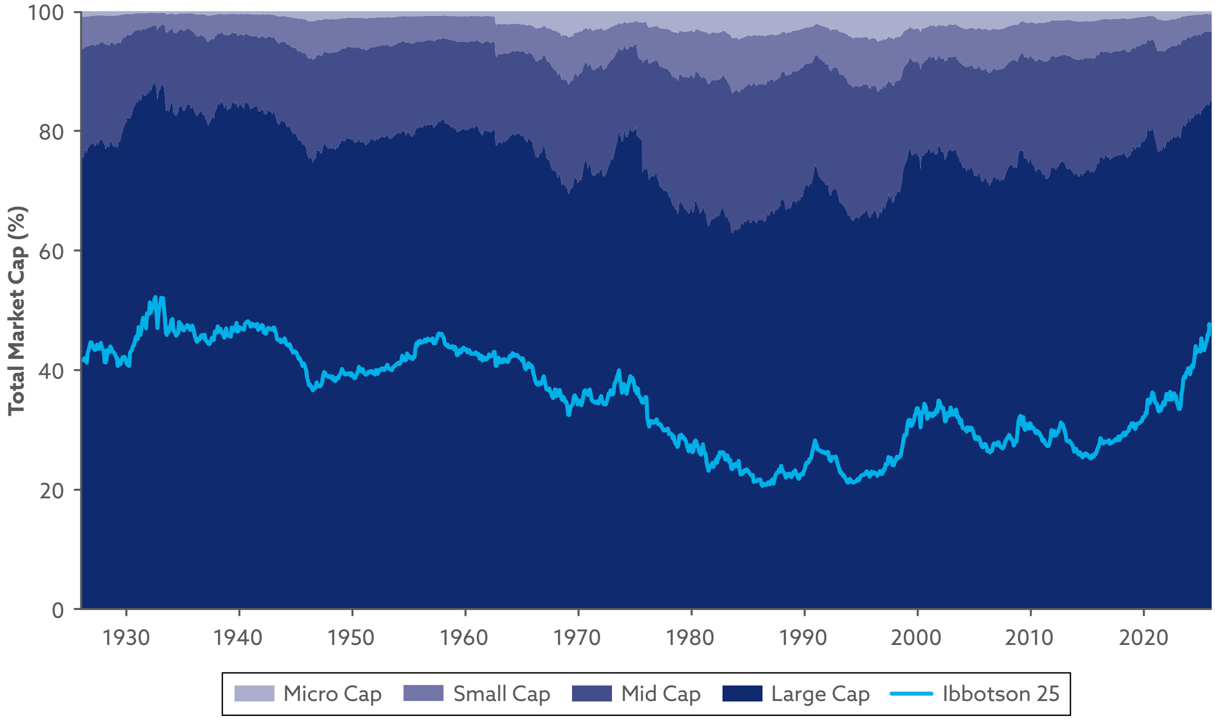
This recent concentration reflects the extraordinary rise of a small number of large technology firms. Such concentration is not without historical precedent: Railroads dominated US market capitalization in the early twentieth century much as software, semiconductors, and artificial intelligence do today. What distinguishes the current episode is that after controlling for the overall breadth of the listed universe, today's concentration is the highest it has been in the entire past century. **Exhibit 4.7** plots industry-level market-capitalization shares using the Fama–French 12-industry classification, showing that technology now accounts for nearly 40% of total market value—a level of industry concentration without parallel in the past century.

To further quantify this effect more rigorously, we use the Herfindahl–Hirschman Index (HHI), a measure of portfolio concentration defined as the sum of squared portfolio weights:

$$\mathrm{HHI} = \sum_{i=1}^{N} w_i^2,$$

where w_i is the market-capitalization share of firm i. Higher values indicate greater concentration. For reference, an equally weighted portfolio of N stocks has an HHI of $1/N$, while a

Exhibit 4.6. Evolution of the Market-Capitalization Share of Each Base Equity Index, 1926–2025



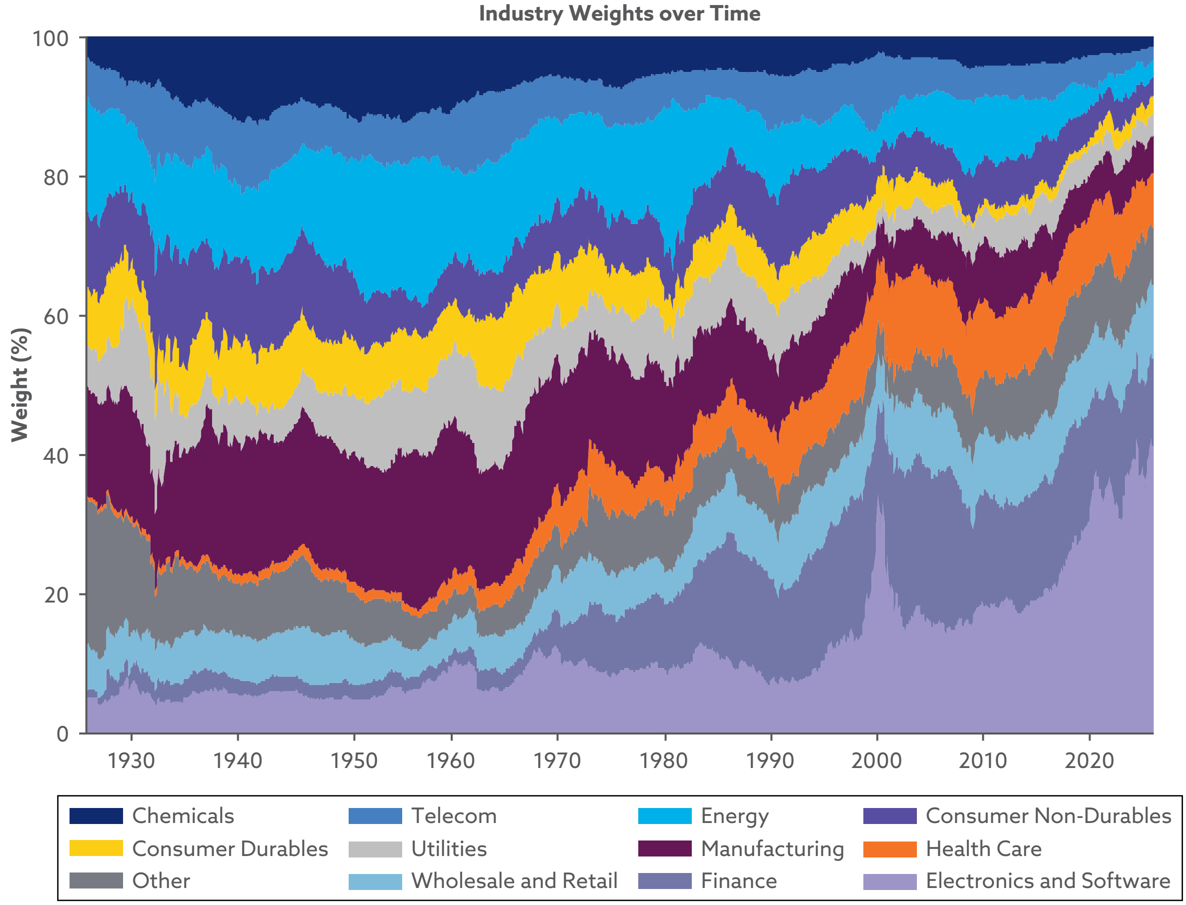
Source: Ibbotson Equity Indices, underlying data from CRSP. Calculations by the authors.

single-firm market achieves the maximum value of 1. We multiply all values by 10,000 for interpretability, so that typical US market values range from roughly 25 to 250 and the maximum attainable value becomes 10,000. In a portfolio context, 1/HHI has a natural interpretation as the effective number of independent positions—the number of equal-sized holdings that would produce the same level of concentration as the actual portfolio. This attribute makes the HHI a direct and intuitive measure of diversification.

Although the HHI is an intuitive and clear way to express concentration, comparing portfolios with different numbers of stocks in them can be difficult. A market with 500 firms will mechanically exhibit a higher raw HHI than one with 5,000, even if the largest companies hold identical relative weight. Thus, to be able to compare the early part of our sample, when the number of listed companies was well below 1,000, with the late 1990s, when publicly listed companies surpassed 7,000, we must apply an adjustment to the HHI. To isolate the actual changes in concentration from changes in market breadth, we scale the raw index by *N*/10,000. This normalization allows for meaningful comparison across decades with varying levels of market breadth.

Exhibit 4.8 plots both HHI series for the entire US market. The raw HHI mirrors the patterns in Exhibit 4.6, with the highest values at the beginning and end of the sample. The adjusted HHI, however, tells a more striking story. After we account for the number of listed firms, today's

Exhibit 4.7. Evolution of the Market-Capitalization Share of the 12 Fama–French Industries, 1926–2025

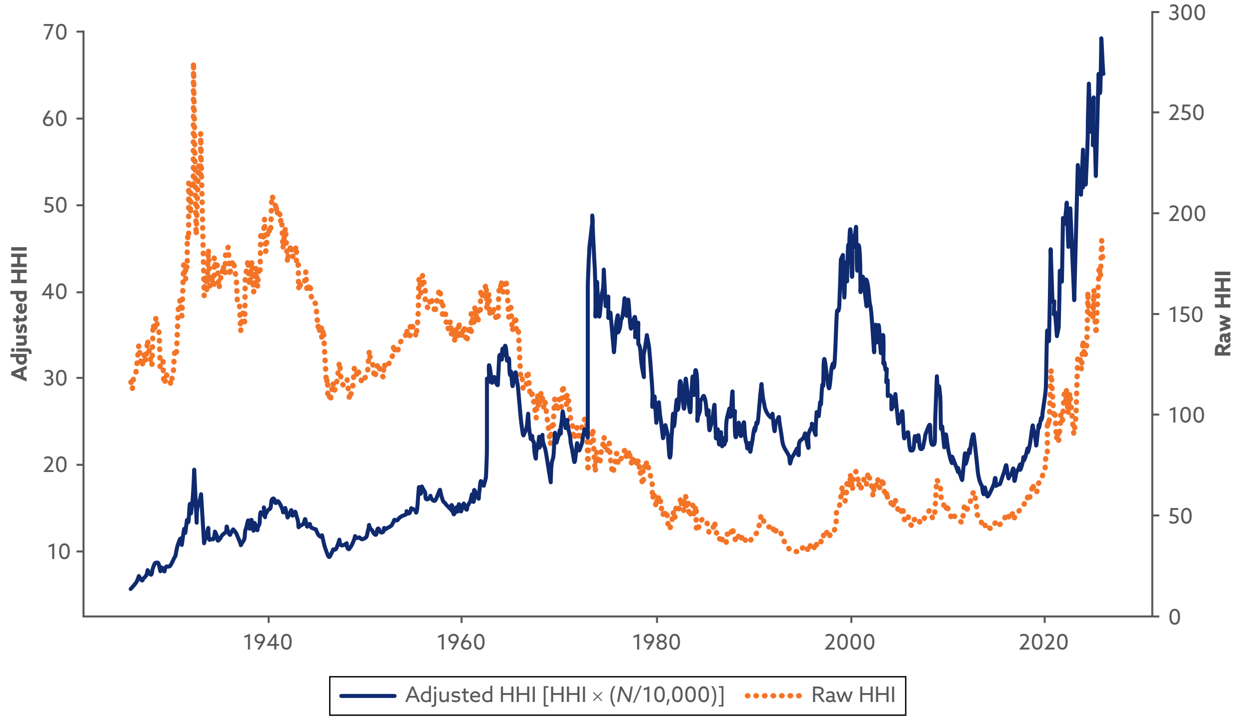


Source: CRSP. Industry labels follow the Fama–French SIC-based 12 Industry portfolio definitions from Kenneth French's Data Library.

concentration is the highest in the entire past century—exceeding even the elevated levels of the 1930s. This is not just a consequence of the decreasing number of firms in the universe from the peak of the late 1990s; instead, it reflects a genuine accumulation of market weight among a very small number of companies.

The practical implication for investors is direct. In the 1975–2000 period, when market weights were more evenly distributed across a rapidly expanding universe of listed firms, a broad cap-weighted index provided more diversification across industries and firm sizes than it does now. Today, that same index delivers a portfolio with a notable tilt toward technology—even more so to the few very large companies in the sector. Because of the strong past performance of this small market segment, holders of the US market portfolio have seen very good returns over the last few years while becoming even more sensitive to the future performance of these stocks. The effective number of independent positions, measured by 1/HHI (10,000/HHI in our case

Exhibit 4.8. Herfindahl–Hirschman Index of the Market Concentration, 1926–2025



Source: Ibbotson Equity Indices, underlying data from CRSP.

after the scaling), has declined sharply in recent years even as the headline number of index constituents remains large. Still, even at current concentration levels, a broad US equity index remains far more diversified than any individual sector or single-stock position.

The Shrinking Small-Cap Premium

The summary statistics presented earlier hinted at a well-documented phenomenon: the small-cap premium. Despite being present over the full century, the premium has diminished substantially in the second half of the sample. In the first half (1926–1975), the arithmetic mean return spread between large- and small-cap portfolios was approximately 3.5 percentage points annually. In the second half (1976–2025), the spread narrowed to around 1.1 percentage points, and for micro-cap stocks, the geometric return actually fell below that of the large caps.

The timing of the decline is worth noting. Banz published his documentation of the size effect in 1981, and there is evidence that the premium began shrinking shortly thereafter. McLean and Pontiff (2016) show systematically that anomaly returns decay after academic publication as capital flows toward documented mispricings and arbitrages them away. The size premium fits this description well: Institutional ownership of small-cap stocks increased substantially through the 1980s and 1990s, improving liquidity as dedicated small-cap investment vehicles became more common. These developments alleviated some of the illiquidity and information risk that had caused investors to previously require a premium as compensation for holding small-cap stocks, and as that underlying risk diminished, so too did the premium it commanded. Fama and French (1993) originally interpreted the size premium as compensation for systematic risk not captured by the CAPM, and under that interpretation, the decline in the premium is consistent with a true reduction in underlying risk rather than a simple correction of a mispricing.

The historical outperformance of small caps that does exist in the full-sample record is moreover heavily concentrated in a small number of exceptional episodes. The years immediately following the Great Depression and the period of high inflation during 1975–1983 account for a disproportionate share of the cumulative small-cap premium. Outside of these episodes, the performance advantage of small-cap stocks over the full century is more modest than the headline numbers suggest. This episodic nature of the premium means that realizing the small-cap premium in practice requires patience and conviction to remain invested through extended periods of underperformance, with no guarantee that the historical episodes of outperformance will repeat.

Despite the decline in the return premium, small-cap and microcap stocks retain diversification value within a broader portfolio. Their return patterns differ sufficiently from large-cap stocks that inclusion in a diversified portfolio can improve the overall risk–return profile even when the standalone expected return advantage of the small caps is modest. Investors should therefore evaluate the benefits of small caps in a portfolio context rather than solely on the basis of their expected excess return.

Decline in Aggregate Volatility

The summary statistics reveal a second structural shift that is easy to overlook: Aggregate equity volatility has declined substantially across all size segments between the first and second halves of the sample. Annualized volatility for large-cap stocks fell from roughly 21% in the 1926–1975 period to approximately 17% in the 1976–2025 period. The pattern holds across all size categories, although the absolute level of volatility remains considerably higher for smaller-cap portfolios throughout.

This long-run downward trend in aggregate volatility is well documented. Comin and Philippon (2005) document the decline in aggregate volatility in the postwar period, attributing it in part to improved macroeconomic stabilization and the reduced share of cyclical industries in the economy. An important nuance is that the decline in aggregate volatility has occurred alongside an increase in firm-level idiosyncratic volatility (Campbell, Lettau, Malkiel, and Xu 2001). This divergence reflects the growing importance of firm-specific information in driving individual stock returns, a development consistent with more informationally efficient markets. Simultaneously, the benefits of diversification have, if anything, increased over time.

The concentration dynamics documented in the previous market concentration discussion add a further layer of complexity to this picture. The declining aggregate volatility of the broad market index partly reflects the growing dominance of large, mature technology firms with relatively stable cash flows, whose increasing index weight mechanically dampens measured portfolio volatility. This compositional effect is distinct from any genuine reduction in the underlying riskiness of the economy, and investors should be cautious about extrapolating the lower volatility of the recent past as a reliable guide to future risk.

Despite the secular decline, the market has remained prone to sharp, systemic crises. The global financial crisis of 2007–2009, the COVID-19 shock of 2020, and the inflation-driven selloff of 2022 each produced drawdowns that, although shorter lived than the Great Depression, were comparable in their immediate severity to many historical crises. Stock and Watson (2002) coined the term "Great Moderation" to describe the post-1980s reduction in macroeconomic volatility, and Blanchard and Simon (2001) documented the same phenomenon in output volatility more broadly. Although the reduction in volatility has been the trend, both sets of researchers were careful to note that reduced average volatility does not eliminate tail risk. For long-run investors, the practical implication is the same: Lower average volatility does not mean lower worst-case volatility, and scenario planning should not be anchored purely to the recent historical distribution.

Annual Returns on the Ibbotson Equity Indices

Although graphs show the big picture vividly, many readers are interested in the year-by-year returns on the indices, so we present them in **Exhibit 4.9**. The exhibit makes it easy to compare returns across size portfolios, and the many booms and busts that occurred over the past century can be studied numerically. The returns in the exhibit are nominal, and year-by-year inflation rates appear in the rightmost column for reference. (The real return, not explicitly shown in the exhibit, is approximately the nominal return minus the inflation rate.)

Exhibit 4.9. Annual Returns on the Ibbotson Equity Indices, 1926–2025

Year	Ibbotson 25 (%)	Large Cap (%)	Mid Cap (%)	Small Cap (%)	Micro Cap (%)	Inflation (%)
1926	16.02	12.51	1.48	–0.92	–6.72	–1.12
1927	34.81	32.72	34.68	27.80	29.04	–2.26
1928	38.71	38.77	39.85	33.22	44.03	–1.16
1929	–9.63	–10.03	–27.09	–38.09	–50.29	0.58
1930	–22.32	–27.02	–36.08	–39.66	–45.69	–6.40
1931	–41.21	–43.33	–46.82	–50.12	–49.65	–9.32
1932	–13.99	–8.79	–6.17	–2.51	8.24	–10.27
1933	46.48	49.93	103.76	115.75	203.47	0.76
1934	–0.58	1.43	10.91	19.80	31.60	1.52
1935	40.10	44.39	43.75	60.94	63.26	2.99
1936	31.14	30.77	36.16	53.18	75.84	1.45
1937	–30.49	–32.59	–41.96	–49.00	–52.34	2.86
1938	23.23	26.58	37.92	43.18	22.02	–2.78
1939	4.76	3.21	–1.47	4.57	0.05	0.00

(*continued*)

Exhibit 4.9. Annual Returns on the Ibbotson Equity Indices, 1926–2025 (*continued*)

Year	Ibbotson 25 (%)	Large Cap (%)	Mid Cap (%)	Small Cap (%)	Micro Cap (%)	Inflation (%)
1940	−6.07	−7.47	−5.11	−5.22	−12.96	0.71
1941	−11.34	−10.75	−8.16	−10.24	−14.19	9.93
1942	14.59	15.19	21.22	26.44	49.76	9.03
1943	22.38	25.04	37.95	58.15	101.72	2.96
1944	15.43	18.23	30.37	41.66	63.23	2.30
1945	26.46	33.02	56.50	64.74	83.17	2.25
1946	−7.15	−5.35	−8.51	−11.23	−13.36	18.13
1947	8.89	4.58	1.19	−2.90	−2.81	8.84
1948	4.46	2.82	0.12	−4.35	−6.59	2.99
1949	18.49	19.71	22.45	21.23	21.60	−2.07
1950	32.40	29.38	30.30	36.70	45.70	5.93
1951	24.23	21.82	18.39	15.30	9.57	6.00
1952	14.94	14.10	11.89	9.64	6.27	0.75
1953	2.63	1.29	−0.85	−2.94	−5.62	0.75
1954	51.98	48.59	56.12	57.48	63.73	−0.74
1955	32.89	26.89	18.62	20.81	22.40	0.37
1956	9.45	8.64	7.76	6.51	2.94	2.99
1957	−9.89	−9.14	−12.62	−17.77	−15.26	2.90
1958	40.42	42.15	56.23	61.87	70.53	1.76
1959	13.38	12.00	15.06	17.40	18.25	1.73
1960	−0.63	1.34	2.20	−3.30	−4.98	1.36
1961	28.71	26.46	28.81	29.53	31.21	0.67
1962	−6.39	−8.93	−13.13	−16.76	−16.52	1.33
1963	24.03	22.30	16.00	18.55	11.99	1.64
1964	16.82	15.70	18.07	16.53	18.56	0.97
1965	7.18	10.68	25.92	35.15	37.98	1.92
1966	−12.03	−9.47	−5.85	−7.15	−8.20	3.46
1967	23.45	21.75	39.94	63.88	103.43	3.04
1968	6.29	9.18	21.08	31.82	50.14	4.72
1969	−2.77	−7.26	−14.69	−22.16	−32.36	6.20
1970	4.55	2.23	−2.01	−9.87	−16.81	5.57
1971	15.34	14.54	21.23	20.32	17.67	3.27
1972	24.35	20.38	9.06	5.58	−1.38	3.41
1973	−15.05	−14.46	−25.94	−34.35	−40.80	8.71
1974	−29.65	−27.47	−25.13	−25.87	−26.72	12.34

(*continued*)

Exhibit 4.9. Annual Returns on the Ibbotson Equity Indices, 1926–2025 (*continued*)

Year	Ibbotson 25 (%)	Large Cap (%)	Mid Cap (%)	Small Cap (%)	Micro Cap (%)	Inflation (%)
1975	31.52	33.98	57.09	60.92	71.52	6.94
1976	20.77	22.06	38.26	46.84	48.19	4.86
1977	–7.87	–7.96	4.25	17.16	28.37	6.70
1978	8.25	5.65	10.59	18.21	26.21	9.02
1979	9.53	17.81	33.60	40.55	43.87	13.29
1980	35.39	32.85	32.87	36.81	39.17	12.52
1981	–11.54	–6.58	4.03	3.60	–2.69	8.92
1982	21.47	19.26	24.89	25.18	25.87	3.83
1983	19.80	20.33	24.65	30.05	33.05	3.79
1984	11.23	7.79	–0.22	–3.41	–11.75	3.95
1985	27.50	33.03	31.06	35.09	28.18	3.80
1986	13.52	18.07	16.33	6.87	3.89	1.10
1987	6.47	3.72	0.73	–6.79	–14.10	4.43
1988	12.91	16.15	21.17	24.88	19.74	4.42
1989	33.58	32.40	23.95	18.83	8.38	4.65
1990	5.03	–2.80	–10.77	–18.24	–27.77	6.11
1991	30.65	31.23	43.22	46.39	49.97	3.06
1992	–0.62	6.73	16.12	17.95	20.80	2.90
1993	2.98	8.06	16.74	18.24	18.67	2.75
1994	4.62	1.38	–3.31	–1.73	–4.19	2.67
1995	40.95	39.11	33.68	28.65	34.20	2.54
1996	27.68	23.36	18.03	15.36	18.13	3.32
1997	37.18	34.52	25.90	24.33	22.91	1.70
1998	41.48	33.43	8.95	0.44	–3.10	1.61
1999	27.33	22.90	29.87	35.14	33.95	2.68
2000	–23.67	–12.69	–4.36	–12.27	–14.30	3.39
2001	–11.85	–14.73	–4.04	7.73	17.95	1.55
2002	–23.36	–21.60	–19.13	–22.36	–16.81	2.38
2003	20.54	27.38	40.58	51.04	73.21	1.88
2004	4.80	9.70	17.58	21.03	15.77	3.26
2005	–0.09	4.76	12.16	6.63	2.17	3.42
2006	17.92	15.74	13.71	16.92	18.05	2.54
2007	8.44	6.85	6.00	–0.72	–8.34	4.08
2008	–32.72	–35.79	–40.18	–37.43	–42.45	0.09
2009	15.70	24.48	41.20	43.89	55.58	2.72

(*continued*)

Exhibit 4.9. Annual Returns on the Ibbotson Equity Indices, 1926–2025 (*continued*)

Year	Ibbotson 25 (%)	Large Cap (%)	Mid Cap (%)	Small Cap (%)	Micro Cap (%)	Inflation (%)
2010	9.99	14.09	28.32	29.27	26.20	1.50
2011	4.18	1.72	−0.22	−3.60	−13.23	2.96
2012	12.85	15.71	17.27	17.21	17.37	1.74
2013	23.74	33.18	39.07	42.39	48.24	1.50
2014	9.17	13.09	9.11	3.88	1.74	0.76
2015	3.44	0.93	−3.69	−6.71	−9.97	0.73
2016	12.48	12.39	15.92	21.51	22.35	2.07
2017	24.17	22.90	20.13	16.03	13.68	2.11
2018	−0.35	−3.62	−9.25	−11.52	−15.23	1.91
2019	34.29	31.19	28.75	23.64	19.87	2.29
2020	29.56	23.99	21.14	22.61	32.92	1.36
2021	29.96	26.45	16.57	12.78	10.77	7.04
2022	−27.14	−19.84	−18.43	−22.07	−26.60	6.45
2023	42.06	29.04	16.90	19.87	2.68	3.35
2024	41.09	27.77	14.57	9.86	10.80	2.89
2025	21.32	19.07	9.53	9.84	13.97	2.68

Source: Ibbotson Equity Indices, underlying data from CRSP and the US Bureau of Labor Statistics.

Why Long-Run Data?

Having documented the long-run behavior of US equities across size segments, we now turn to two more fundamental questions: How much statistical confidence should investors place in these historical estimates, and what does a century of data actually buy in terms of precision? At first glance, long-run data—especially data from nearly a century ago—may seem of limited value for decisions about the future. The US economy today is different from the economy of decades past. The drivers of economic growth, inflation, and market structure have all changed. Why, then, should distant history inform our beliefs about future risk and return?

First and foremost, history is a record of important events and their causes and consequences. History is particularly useful in evaluating rare, major economic events, precisely because they happen so infrequently. For example, in recent years, history has been an indispensable reference point for assessing the effects of a pandemic, inflation shocks, radical technological change, and disruptions in global trade. History may not directly predict what will happen, but it provides an analytical framework and guidance about the factors we should consider in modeling future consequences.

The virtue of continuous, long-term market data, as opposed to simply historical records of rare events, is the large-sample statistical properties of such data. There are important caveats to relying on even a large sample of data. Most importantly, the data exist for a reason. In the case of a century of US equity returns, the data exist because the US market survived (and thrived) for a century, unlike some other major markets that were interrupted or shut down by war, revolution, or other circumstances in the twentieth century. Accounting for this potential bias in the data is important.

Another thing to keep in mind is that the data we have today are bounded by history. The tails do not contain the limits of future extreme outcomes. The one-day crash in 1987 was roughly double the magnitude of the biggest one-day crash in 1929. An analyst in 1986 looking back at history might have thought such a crash impossible. Someday, there will almost certainly be another crash that sets a new record. Do not trust historical data as the only set of potential future outcomes.

With these caveats in mind, in this section, we explore how having lots of data can help assess the probabilities of future outcomes and provide higher levels of confidence. And we will show how "lots" can have different meanings for different measures.

Continuous, long-term historical data yield desirable statistical properties when estimating the key parameters investors care about: expected returns, volatility, and correlations. The central insight is that these three quantities respond to different dimensions of the data. As we will show, more calendar time helps pin down expected returns, whereas more frequent observations help pin down volatility and correlation. Understanding this distinction clarifies exactly what a long dataset buys—and what it does not.

Estimation of Expected Returns: Time Span Matters

The most important question in the practice of long-run investing is, How precisely can we estimate expected returns? One way to form an expectation of the future is to use historical data and estimate the mean return. This statistic and the accuracy of the estimate depend entirely on the length of the time period covered by the data, not on how frequently returns are measured within it.[3]

Intuitively, expected return is a slow-moving signal buried in noise. Adding more observations by increasing the frequency—for example, by switching from monthly to daily data over the same time window—does not help, because each additional observation carries proportionally less information.[4] What actually tightens the estimate is extending the sample window itself.

Formally, the standard error, SE, of the estimated mean return, $\hat{\mu}$, is

$$\mathrm{SE}(\hat{\mu}) \approx \frac{\hat{\sigma}}{\sqrt{T}}, \tag{4.1}$$

where $\hat{\sigma}$ is the estimated annual volatility and T is the number of years in the sample. To halve the standard error, one must quadruple the sample length. Thus, going from a 25-year sample

---

[3] The estimates presented assume that both $\hat{\mu}$ and $\hat{\sigma}$ are constant over the sample and not serially correlated.

[4] Each return observation contains a signal proportional to the length of the period and noise proportional to its square root. Shrinking the interval and moving to higher-frequency data make both the signal and the noise smaller, but the noise shrinks more slowly than the signal. These effects end up canceling out, so adding more observations by increasing the frequency never improves the precision for the estimate of the mean. Only extending the sample window allows the signal to grow relative to the noise.

to a 100-year sample cuts the uncertainty in half. This dynamic illustrates both the value of a century-long dataset and the practical difficulty of achieving high precision in expected return estimation.

The takeaway is that for expected returns, what matters is how long the sample runs, not how finely it is measured.

Estimation of Volatility and Correlation: Frequency Matters

In addition to expected return, investors also care about risk, especially volatility and correlation. The estimates of volatility and correlation behave differently from the mean. Here, the precision improves with the number of observations, n, not the time span. Thus, higher-frequency data, such as daily rather than monthly, can in fact help with risk measures, at least to the point at which microstructure noise (e.g., bid–ask bounce) begins to add noise to the signal.

The standard error of the estimated volatility, the standard deviation, shrinks as n increases:

$$\mathrm{SE}(\hat{\sigma}) \approx \hat{\sigma}\sqrt{\frac{1}{2n}}. \tag{4.2}$$

The same logic applies to correlation. Volatility is estimated by averaging squared returns, and correlation is estimated by averaging cross-products of returns. Each additional observation contributes a fresh squared or cross-product term, whereas for the mean, intermediate observations telescope away and add nothing. This is why finer sampling (to an extent) does reduce estimation error for risk parameters. Therefore, a practitioner using daily data has roughly $\sqrt{252} \approx 16$ times more precision regarding volatility than one using annual data over the same time window.

For this reason, a long dataset and a high-frequency dataset are complements, not substitutes, because they solve different problems. A century of annual data is invaluable for expected return estimation but nearly useless for improving volatility precision beyond what a decade of daily data already provides. Conversely, switching from annual to daily observations over the same 10-year window dramatically sharpens risk estimates but leaves the expected return estimate essentially unchanged.

To summarize, for volatility and correlation, precision depends primarily on the number of observations. Daily data can provide far more precision than annual data over the same window.

Summary

This chapter introduces the Ibbotson Equity Indices, a new family of capitalization-weighted US equity portfolios constructed using security-level data from CRSP spanning 1926–2025. The indices are designed to provide a transparent and economically meaningful representation of the US public equity markets and various size segments over a full century.

The construction methodology emphasizes stability over time. By relying on CRSP data and by using consistent size definitions—CRSP deciles before 1975 and fixed numeric rank market-cap thresholds thereafter—the Equity Indices allow for comparisons across decades that reflect economic change in the US markets. All the Equity Indices are constructed using full market-cap weighting, a decision that preserves continuity across the entire sample, while tying the index weights to the actual economic footprint of publicly traded firms.

Using the Equity Indices' coverage of the past century, we documented well-known facts about the US equity markets. First, long-run compounded average returns have been remarkably strong, approximately 10% per year on average, with meaningful differences across size segments. Second, although small-cap stocks have historically earned a premium over their large-cap peers, the magnitude of this premium has declined in the second half of the sample. Third, aggregate volatility has declined over time, despite the occurrence of several rare systemic crises during the past 50 years.

The evolution of market concentration provides additional perspective. Capitalization weighting naturally reflects changes in the distribution of firm size. Periods of elevated concentration in the megacap segment—for example, in the 1930s or currently—reduce the effective breadth of diversification in broad market indices, although they remain well diversified when compared to smaller portfolios.

Finally, we demonstrated why long-run data remain valuable for investment analysis. In particular, the precision of estimation of the sample mean depends on the total span of the sample, improving at a rate proportional to $1/\sqrt{T}$, while volatility and correlation estimates improve with the number of observations. A long, high-quality dataset can therefore enhance statistical confidence while allowing investors to focus on subperiods they believe are most relevant to current market conditions.

Taken together, the Ibbotson Equity Indices provide a consistent empirical foundation for understanding US equity market behavior through different market conditions and size segments. Thus, they are intended to serve as a durable benchmark for research and supplement long-term investment decision making.

Key Takeaways

- The Ibbotson Equity Indices are built from CRSP security-level data using stable size definitions and full market-cap weighting, ensuring as much methodological consistency as can practically be achieved, over the 1926–2025 period.
- The broad US equity market generated approximately 10% annual geometric returns over the past century, with persistent but time-varying differences across size segments.
- Equity market volatility has declined across all size portfolios over the past century.
- The capitalization of the top 25 firms as a share of the entire market has fluctuated between roughly 20% and 50%. When this share is high, the effective diversification of cap-weighted indices is reduced, even though these indices remain broadly diversified overall.
- Long-run data provide greater precision when estimating the sample mean, and the precision improves at the rate $1/\sqrt{T}$. The longer the sample, the better the estimate.

References

Banz, Rolf W. 1981. "The Relationship Between Return and Market Value of Common Stocks." *Journal of Financial Economics* 9 (1): 3–18. doi:10.1016/0304-405X(81)90018-0.

Blanchard, Olivier, and John Simon. 2001. "The Long and Large Decline in US Output Volatility." *Brookings Papers on Economic Activity* (1): 135–74. doi:10.1353/eca.2001.0013.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu. 2001. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk." *Journal of Finance* 56 (1): 1–43. doi:10.1111/0022-1082.00318.

Comin, Diego, and Thomas Philippon. 2005. "The Rise in Firm-Level Volatility: Causes and Consequences." *NBER Macroeconomics Annual* 20: 167–201. doi:10.1086/ma.20.3585419.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2002. *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton, NJ: Princeton University Press. doi:10.1515/9781400829477.

Doidge, Craig, G. Andrew Karolyi, and René M. Stulz. 2017. "The US Listing Gap." *Journal of Financial Economics* 123 (3): 464–87. doi:10.1016/j.jfineco.2016.12.002.

Fama, Eugene F., and Kenneth R. French. 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.

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.

McLean, R. David, and Jeffrey Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *Journal of Finance* 71 (1): 5–32. doi:10.1111/jofi.12365.

Reinganum, Marc R. 1981. "Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings' Yields and Market Values." *Journal of Financial Economics* 9 (1): 19–46. doi:10.1016/0304-405X(81)90019-2.

Stock, James H., and Mark W. Watson. 2002. "Has the Business Cycle Changed and Why?" *NBER Macroeconomics Annual* 17: 159–218. doi:10.1086/ma.17.3585284.

CHAPTER 5. STATISTICAL ANALYSIS OF EQUITY RETURNS

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

Introduction

In this chapter, we examine the long-run behavior of US equity returns through several complementary lenses. We begin by assessing how performance has evolved over decades and across investment horizons.

Over the past century, US markets have not grown in a smooth or uniform fashion. Different decades have delivered sharply different outcomes. Long stretches of strong performance have sometimes been followed by lost decades, and the experience of an investor in one era can look very different from that of an investor in another period. To capture this variation, we begin by looking at both rolling returns and decade-by-decade returns to highlight the dispersion of realized outcomes and the importance of horizon length.

The deeper story of this chapter, however, concerns what the shape of return distributions reveals about risk. Mean returns and standard deviations—the first two moments of a distribution, discussed in the previous chapter—make up the standard toolkit, but they conceal features that matter enormously to investors. These characteristics are asymmetry, heavy tails, and the frequency of extreme events. Equity returns are not Gaussian (normally or lognormally distributed), and the departure from normality is the driving force behind why crashes happen more often than tidy models predict, why a handful of stocks generate most of the market's long-run wealth, and why diversification is more consequential than it appears when viewed through a variance lens alone. By examining skewness and kurtosis alongside the historical record of drawdowns and rolling returns, we gain a more complete and honest picture of what equity investing has actually looked like.

The time-series perspective on stock returns in this chapter complements the sector and industry analysis in Chapter 4, which goes beyond observing what the market has returned to ask when, how, and from which parts of the economy those returns have come.

Past Performance

Over the past century, US equities have delivered extraordinary compounded growth, but that long-run success tends to mask a highly volatile and uneven path. The equity premium has rewarded investors who are willing to tolerate risk, yet the journey has been marked by deep drawdowns, regime shifts, and extended recoveries. As a result, most US equity investors did not earn the long-run average return, for various reasons (the most important being the subperiods over which they were invested). To understand market volatility and its implications for the past century's returns, we will first inspect return paths within each decade, followed by long-horizon rolling returns. Both perspectives reveal how realized outcomes can differ across decades and other holding periods, which full-sample statistics and wealth indices tend to wash away.

Returns by Calendar Decade

Examining cumulative returns within each calendar decade highlights the booms, busts, and recoveries that have defined the investor experience. **Exhibit 5.1** plots the 10-year return paths for each decade of the past century for the Ibbotson All Cap Stock Index™. Each line represents the growth of $1 invested at the beginning of each decade.

What stands out is that despite numerous major drawdowns, markets have usually recovered fairly quickly. Exceptions are the Great Depression and the early years of World War II, which constituted the largest and longest drawdown; the 1970s stagflation; and the more recent "lost decade" that included the 2000–2002 dot-com collapse and the global financial crisis of 2007–2009. Note that the decade-by-decade plots can understate risk because crises rarely align with calendar decades and severe declines are often split between two panels in the exhibit. Still, this decade analysis gives a useful picture of market volatility, including booms, crashes, and recoveries. All of them are defining features of equity investing.

The volatility observed in these 10-year periods naturally translates to drawdowns—that is, periods when the market falls from a peak and has yet to surpass the old high. **Exhibit 5.2** summarizes the duration and magnitude of the major market-wide US equity drawdowns from 1926 to 2025 using the nominal Ibbotson All Cap stock total return index. Excluding single-month declines, the average drawdown lasted roughly 11 months; including single-month declines reduces the average to roughly 7 months. Fourteen drawdowns exceeded one year in duration. Outside of the 1929 crash, the effects of which persisted for more than a decade, recoveries from major downturns, such as in 1973, 2000, and 2007, occurred within three to six years.

For investors, the duration of a drawdown can matter just as much as the magnitude. A short-lived, sharp decline may be tolerable for an investor who is decades from retirement, but even a mild drawdown may be consequential for someone already retired with a majority of assets held in equities. Thus, a thorough understanding of past drawdowns and their duration is needed when thinking about long-term investing, especially if the investments face some sort of liability side, as is the case with pension funds. For these types of investors, the relevant risk is not only the likelihood of a drawdown but also the time to recovery relative to the investment horizon.

Exhibit 5.1. All Cap Total Returns by Decade with Drawdowns Highlighted, 1926–2025

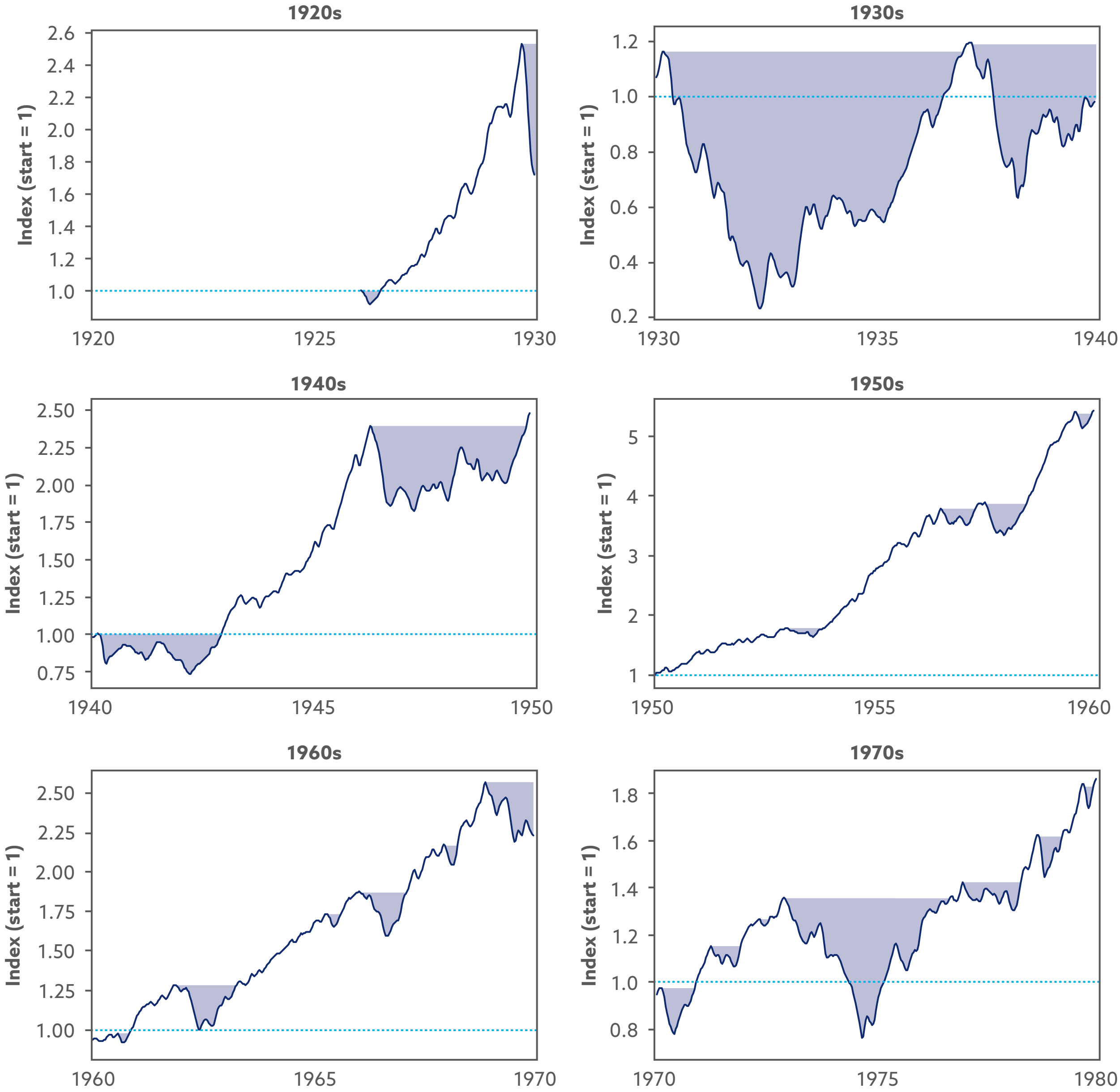
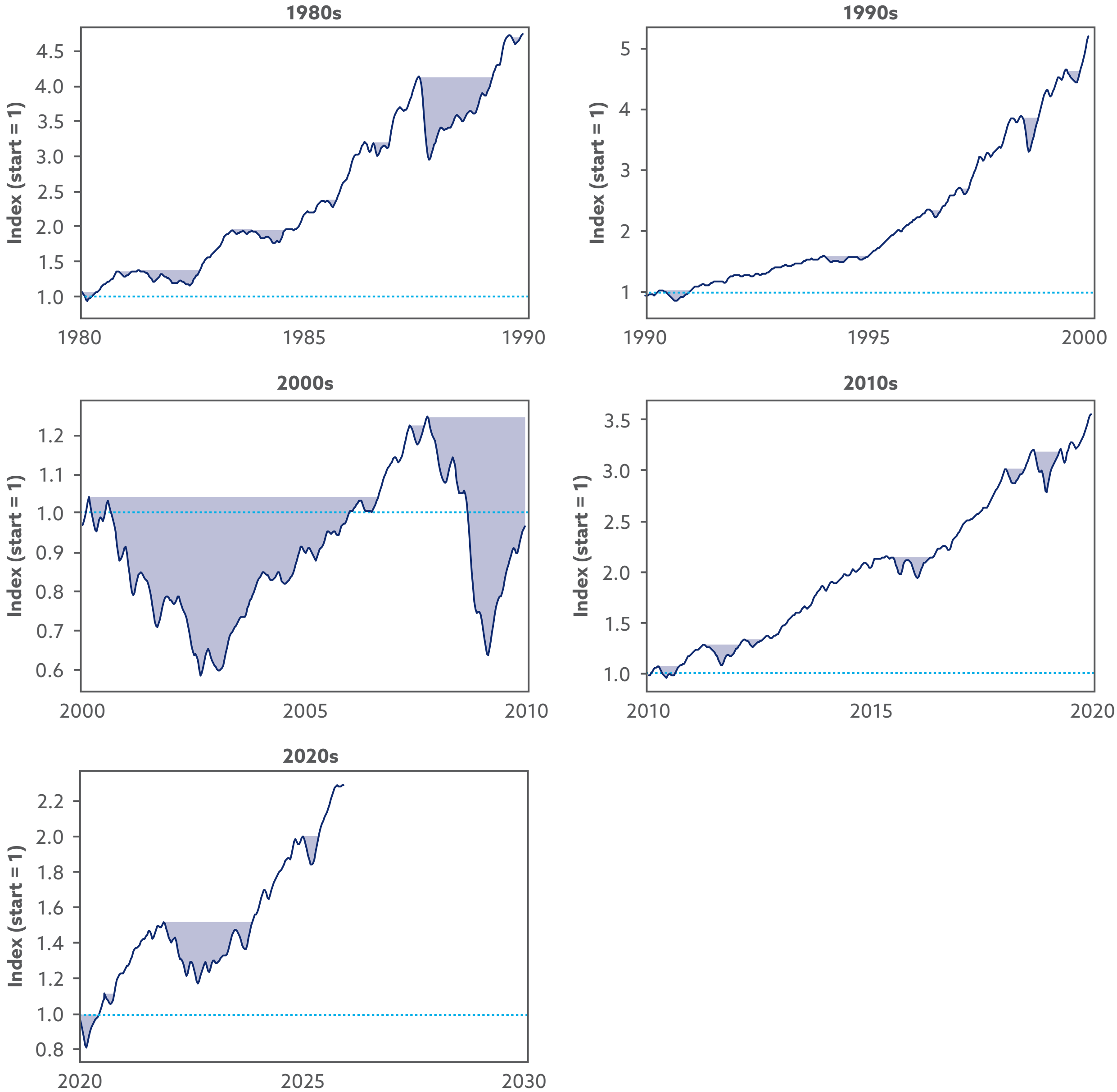
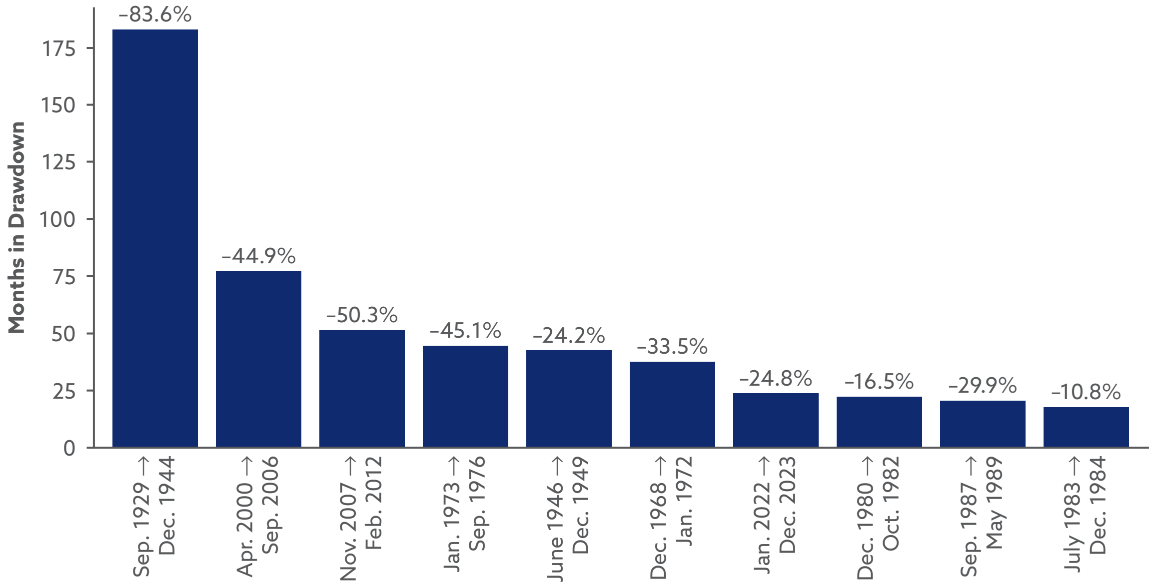


Exhibit 5.1. All Cap Total Returns by Decade with Drawdowns Highlighted, 1926–2025 (*continued*)



Source: Roger G. Ibbotson. Data from the Ibbotson All Cap stock index.

Exhibit 5.2. Histogram of US Market Drawdown Lengths (months), with Maximum Drawdown (%)



Source: Roger G. Ibbotson. Data from the nominal Ibbotson All Cap stocks total return index.

Rolling-Period Returns

To move from decade-by-decade views to a longer-run perspective, **Exhibit 5.3** plots 20-year compound annualized real returns for both US equities and 20-year Treasury bonds. These overlapping windows show how the realized equity premium and the term premium of long-term bonds over bills have fluctuated through time.

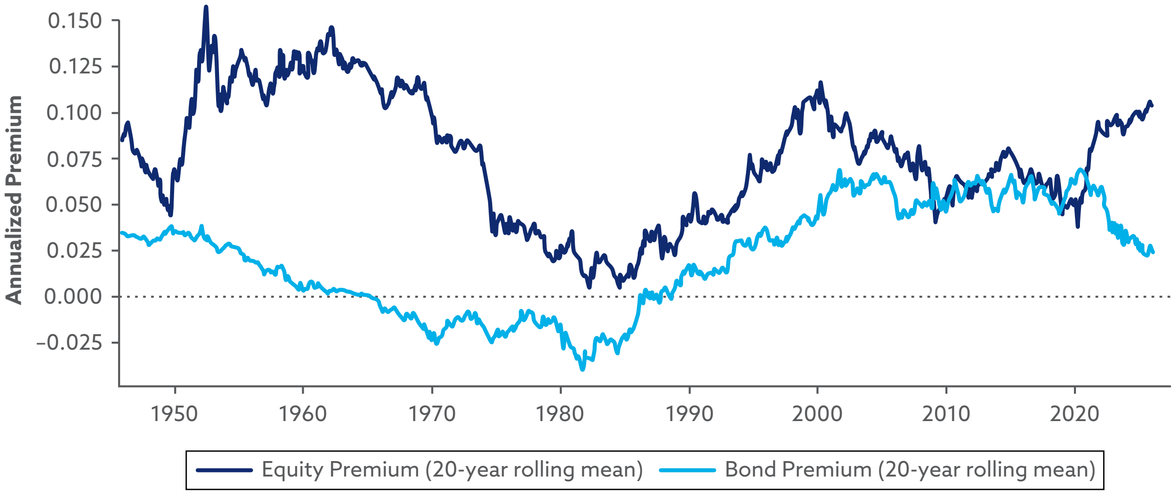
Despite these wild fluctuations, equities have generally outperformed long-term Treasury bonds over the post-1926 period. On average, large-cap US equities have delivered an arithmetic mean real return 6 percentage points higher than that of 20-year bonds,[1] with the two asset classes having Sharpe ratios of 0.45 and 0.18, respectively.

This gap has fueled the "stocks for the long run" narrative (Siegel 2022) because of the size and persistence of the realized equity premium. McQuarrie (2023) questioned the universality of this finding by introducing a new dataset for the pre-1926 period. He showed that prior to 1926, the advantage of equities over bonds varied substantially, with bonds beating stocks for long stretches of time. Our focus here remains on the modern (post-1926) era, where a century of rolling 20-year returns demonstrates both the persistence and the variability of the US equity premium when measured against long-term government bonds.

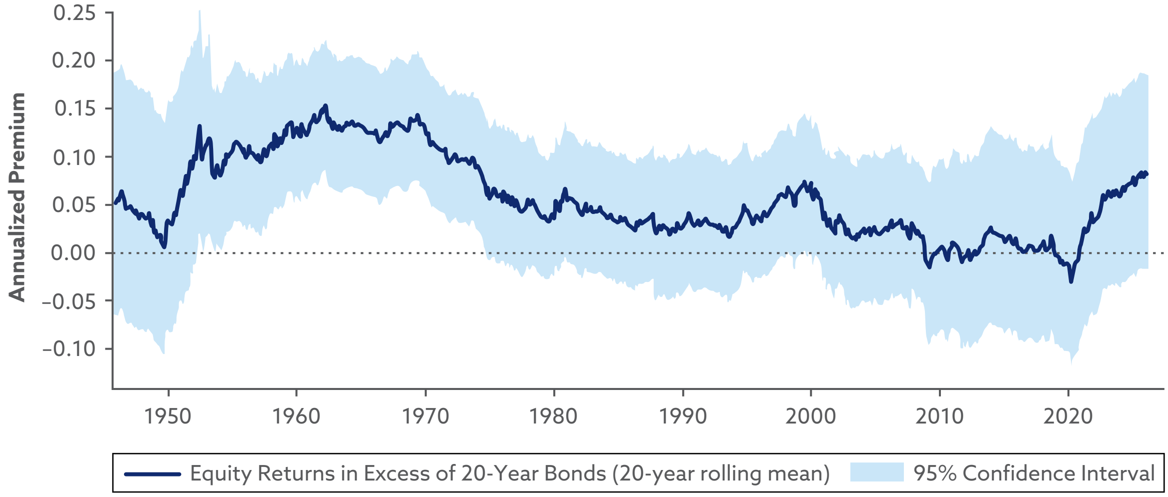
[1] This is the difference between the arithmetic, not geometric, means.

Exhibit 5.3. 20-Year Rolling Premia of Large-Cap Equities and 20-Year Treasury Bonds over T-Bills, 1946–2025

A. Stock and Bond Premia over T-Bills



B. 20-Year Rolling Premium of Stock Return in Excess of Bond Return, with Confidence Bands



Source: Roger G. Ibbotson. Data from Ibbotson Indices.

Note: Confidence bands are based on Heteroskedasticity and Autocorrelation Consistent Newey–West standard errors.

These patterns underscore a recurring theme: The realized history of equity returns is punctuated by events whose frequency and magnitude are difficult to square with simple distributional models. The decade-by-decade and rolling return views document what happened. To understand why such events occur—and how often investors should expect them to recur—we need to look more carefully at the full shape of the return distribution, not just its center.

Distributions of Stock Returns

Every numerical variable, from human height to exam scores, follows some distribution, a pattern that describes how often different values occur. Many natural phenomena tend to cluster around an average with fewer observations far from the center. Empirical distributions can be calculated from data. Theoretical distributions are derived from mathematical models of the data-generating process. The most familiar of these is the bell-shaped, Gaussian normal distribution, which is symmetric around the mean and has approximately two-thirds of the occurrences within one standard deviation of the mean.

Are Stock Returns Best Modeled as Normal, Lognormal, or Something Else?

The starting point is the distribution of returns over a single period—say, a month or a day. Neither the normal nor the lognormal distribution is a perfect fit, but the lognormal is a better approximation. The primary reason is limited liability: A long-only equity position cannot lose more than 100%, which rules out the symmetric unbounded losses implied by the normal distribution. Even so, empirical single-period returns exhibit fat tails and skewness that the lognormal does not fully capture, a phenomenon on which we focus in this chapter.

A related but distinct question concerns the behavior of compounded wealth over long horizons. Because compounded wealth is the product of successive gross returns, taking logarithms converts it into a sum of log returns. This sum then tends toward normality as the number of periods grows because of the central limit theorem, implying that long-horizon wealth tends toward lognormality even when individual-period returns are not themselves lognormal. Chapters 19 and 20 explore the forecasting implications of this result.

This distinct property of periodic return distributions is important and has a long history of research in financial economics. For example, in a seminal option pricing formula, Louis Bachelier (1900) assumed a normal distribution for bond returns, which implied the possibility of negative prices. Maurice Kendall (1953) made a similar assumption in an early test of the random walk hypothesis. M. F. M. Osborne (1959) was the first to recognize that this assumption violated the lower bound of –100% return and to propose that stock returns are lognormally distributed.

Benoit Mandelbrot (1963) raised an important challenge to the lognormal assumption for asset returns. He observed empirically that they had "fat tails"—in other words, that extreme high and low log returns were more common than predicted by the model. He hypothesized that a stable distribution with undefined variance was a better fit.[2] Because the foundation of modern portfolio theory is the tradeoff between measurable risk and return, the stable distribution with undefined variance invalidates most of its key implications.[3] Eugene Fama (1963, 1965) was the first to test the stable model against the lognormal with stock market data and firmly rejected lognormality. Since then, the fat-tail feature of asset return distributions has remained both a theoretical challenge and a very practical concern in financial economics. It is popularly known as the "black swan" problem—the notion, popularized and more formally studied by

---

[2] A simple example of a stable distribution with undefined variance is a variable with tail probability $PX > x = x - \alpha$ for $x \geq 1$ and $1 < \alpha \leq 2$.

[3] See, by way of comparison, Kaplan (2012).

Nassim Taleb (2010, 2020), that extreme events happen more frequently than any probabilistic model, even the stable distribution, can predict.

On the theoretical side, a number of resolutions to the issue of fat tails have been proposed since Mandelbrot (1963). Stanley Kon (1984), for example, showed that a mixture of normal distributions can also generate fat-tailed distributions—a solution that fits nicely with a pattern of economic regime changes in which some periods have high risk and low return. Other solutions model the stock-return-generating process with jumps as well as continuous variation. Robert Merton (1976) introduced this approach to calculate the price of options when a continuous hedge to the underlying asset is not possible. Robert Engle (1982) introduced a time-varying, conditional volatility model of stock returns (generalized autoregressive conditional heteroskedasticity, or GARCH) and showed that this process could also generate fat-tailed distributions. Further advances have been made in modeling and empirically estimating the variation in the volatility of tail risk itself.[4] Thus, for most practical financial applications, as long as a variable such as a stock return has a finite bound, there are useful modeling tools to evaluate risk. One approach to modeling that does not depend on regime changes, jumps, or innovations in volatility is to adjust the normal distribution for return asymmetry and allow for fatter tails. In Chapter 19, Paul Kaplan shows how to use the more general Johnson distribution in estimating the probability of investor ruin over long investment horizons.

Why Investors Should Care About the Distribution of Returns

Why does all this matter for investors? It matters because the return-generating process and the resulting shape of the distribution of realized returns tells us not only about the typical fluctuations but also the rare events that have an outsized influence on performance and risk. Knowing whether the return distribution of an asset is symmetric or skewed, thin or fat tailed, can set our expectations about the likelihood of likely extreme gains or losses and how the presence of the asset in a portfolio could impact the degree to which the portfolio is effectively diversified.

The good news for long-term investors is that log returns are a very good first approximation for forecasting the cumulative wealth distribution. As shown in Chapter 3, compound growth is the product of a sequence of single-period growth rates, defined as 1 plus the return. Taking the natural logarithm of these growth rates turns this calculation into a summation of random variables. The central limit theorem shows that the mean of a series of independently and identically distributed random variables drawn from any distribution with a defined variance—not just the normal or lognormal—converges to a normal distribution. This convergence result has been extended to nonindependent and nonidentical distributions.

Advances in modeling the return-generating process have been particularly relevant to shorter-horizon investors as well. For example, in option pricing, the value of derivative contracts depends crucially on understanding the distribution of future prices in the period up to and including expiration of the option. The classic Black–Scholes option pricing model assumes a lognormal distribution of the return to the underlying security. Researchers have consistently shown that this assumption can lead to the mispricing of out-of-the-money options. For example, out-of-the-money put options on the S&P 500, which pay off on big market declines, regularly appear overpriced, either because the perceived probability of a

---

[4] See, for example, Gabaix (2012); Santa-Clara and Yan (2010); Wachter (2013); Bollerslev and Todorov (2011); Bollerslev, Todorov, and Xu (2015); Andersen, Fusari, and Todorov (2020); Calvet and Fisher (1999).

Exhibit 5.4. Probability of Experiencing at Least One Single-Day Crash in the Dow Jones Industrial Average of More Than 10% Within Your Remaining Lifetime

Investment Horizon	Lognormal, i.i.d. Return Distribution	GARCH (1,1) Time-Varying Mean and Volatility Clustering Model	Johnson Distribution[a]	Historical Frequency Model
60	0%	31%	72%	79%
50	0%	27%	65%	73%
40	0%	22%	57%	65%
30	0%	17%	47%	54%
20	0%	12%	35%	41%
10	0%	6%	19%	23%

[a]with estimated mean, standard deviation, skewness, and kurtosis parameters.

stock market crash is higher than the return-generating model implies or because the market demands a higher than theorized compensation for insuring against that risk. Hence, properly modeling tail risk is essential for option trading and hedging.

Planning for Crashes

Another use of the higher moments of distributions is in planning for crashes. The single-day crash of more than 20% in the S&P 500 on 24 October 1987 was an event that had a virtually zero chance of happening if stock returns actually followed a lognormal distribution.[5] In fact, none of the 10 worst single-day crashes in the 140-year history of the Dow Jones Industrial Average of daily stock prices would have had a reasonable chance of occurring—and yet "black swan" events happen. Most readers will recall 16 March 2020, when the Dow dropped by more than 12%—another event that the lognormal model would say is impossible.

To show how the return distribution is relevant in planning for market crashes, we will ask, What are the chances you will experience a single-day loss in a diversified equity portfolio of more than 10% within your remaining lifetime? **Exhibit 5.4** shows probability estimates for this event for various remaining lifespans. The lognormal model says such a crash will not happen. The empirical frequency column on the far right, however, says that it can happen because it did. This column simply counts the number of these events that happened in the history of the Dow, divides by the number of years in that history, and then multiplies by the remaining lifespan. An important caveat is that the historical frequency depends on the fact that the Dow Jones Industrial Average—and the US stock market itself—survived over the past 140 years.

[5]The 19 October 1987 crash had a 1 in 10^{128} chance of occurring under the lognormal assumption. For comparison, the estimated number of stars in the universe is 10^{24}.

Historical frequencies do not take into account the possibility of events that could have happened but did not happen in the observed sample. As such, they are a bounded—and thus likely to be a conservative—estimate of future extreme events. The benefit of using a theoretical model is that it extends the bounds of the empirical distribution, alerts investors to the possibility of extreme events, and provides a probability of them occurring that is not bounded by the limitations of the historical data.

Among the most widely used tools to model time-varying returns is the GARCH model. It is a sophisticated econometric approach that starts with the observations that volatility itself is volatile—that is, some periods are more volatile than others—and that high-volatility periods tend to cluster together. GARCH accounts for these attributes by allowing today's volatility estimate to depend on recent returns and recent volatility. As a result, it produces more realistic probability estimates for extreme events than the simple lognormal model, as shown in Exhibit 5.4. GARCH still underestimates the historical frequency of crashes, however, because its strength lies in short-horizon volatility clustering rather than in capturing the extreme tail behavior that drives the single-day declines.

Unlike the GARCH model, the Johnson model—an econometric model that uses the first four moments of the return distribution, presented in detail in Chapter 19—assumes that stock returns are independently and identically distributed (i.i.d.). Despite this assumption, which is not supported empirically, the model nevertheless captures the effect of variation over time in the mean and variance of the return distribution. It accomplishes this seemingly impossible task by taking into account kurtosis (discussed later), a moment of the distribution that takes on a higher value if volatility has changed over time. The Johnson probability estimates for a 10% one-day decline are reported in the middle column of Exhibit 5.4. This model assigns a higher probability of a crash than does the empirical distribution, suggesting that investors could expect an even higher chance of a single-day crash than history thus far indicates. That result is consistent with our earlier observation that the historical estimate is conservative.

The takeaway from this exercise is that most investors should be prepared to experience another day when the market drops by 10%. For those with a 50- or 60-year investment horizon, such a crash has a 70%–80% probability of occurring. For those with a 10-year horizon, the probability of a 10% drop is roughly 20%–25%. No theoretical model of asset return distribution is perfect, but neither is reliance solely on the realized history of returns. Nevertheless, econometric methods that allow for fatter tails and skewness produce probability estimates that are more likely to be accurate.

Distributions of Returns: Summary

In short, the distribution of asset returns tells the deeper story of risk. Empirical distributions describe what did happen; theoretical distributions describe what could happen and assign probabilities to these outcomes. Both are useful. Theoretical distributions have "tails"—that is, they assign probabilities to extreme events even if such events did not occur historically. However, they are mathematical constructs that are tractable approximations to the underlying processes generating data. They do not and cannot fully capture the economic processes, shocks, transitions, and market evolution that will inevitably characterize the future. Although researchers have long recognized the limitations of the lognormal model of stock returns and have worked to improve on it, particularly for the purposes of asset pricing and risk modeling, the model remains an effective first approximation to estimating long-run investment performance.

Theoretical distributions are useful tools, but empirical distributions based on historical data are the fundamental basis for understanding the return-generating processes underlying asset returns. To describe these distributions with more rigor, statisticians summarize their basic contours using higher-order moments—parameters beyond mean and standard deviation that measure the key dimensions of the shape of the distribution and further help us understand extreme outcomes. We explore these in detail next.

Moments of Distributions

Moments are quantitative measures that describe the shape of a distribution. The first two moments are the mean and the variance, which capture the average return and its volatility. In Chapter 4, we examined the differences in these two properties for different size indices and saw how their estimation can benefit from the length of the sample for the mean and the number of observations for the variance and standard deviation. But distributions can differ in other ways even when they share the same mean and variance. To capture these differences, we turn to the third and fourth moments, known as skewness and kurtosis.

The third and fourth moments reveal information about how returns behave in the tails, describing the extreme positive or negative outcomes that investors care most about. Throughout this subsection, all statistics are based on monthly returns because at this horizon, distributions remain meaningfully nonnormal and returns of this frequency are widely used by investors of all levels.

Skewness: A Measure of Asymmetry in Returns

Skewness measures how tilted a distribution is to one side or the other. For a normal distribution, which is perfectly symmetric, skewness equals zero. If skewness is negative, then the left tail is longer and the bulk of the distribution is concentrated on the right side of the mean. A negatively skewed distribution can also be called left tailed or left skewed. In contrast, a positive skew indicates a longer right tail with the mass concentrated on the left. In a positively skewed distribution, the mean is lower than that of a negatively skewed distribution but there are more extreme positive events because of the long right tail.

Mathematically, skewness is the third moment:

$$\frac{E[(X-\mu)^3]}{\sigma^3}. \tag{5.1}$$

But the intuition is simpler: Skewness tells us on which side of the distribution the extremes or "surprises" tend to occur.

In financial markets, skewness has recognizable patterns. Individual stocks often exhibit positive skewness: Most monthly returns are small, but occasionally a firm's stock surges dramatically following good news, innovations, or other shocks. Index-level returns are more likely to be closer to symmetric because aggregation diversifies away many of these firm-specific extremes. Both very large losses and very large gains become much rarer for broad portfolios than for individual stocks. In broad market indices, skewness is often slightly negative or close to zero, reflecting the fact that large systematic declines and crashes occasionally pull the left tail outward. Smaller-stock indices can, however, retain more of the positive asymmetry from their constituents.

Skewness is particularly important for portfolio choice. Hendrik Bessembinder (2018) and Bessembinder, Chen, Choi, and Wei (2023) show that a small fraction of companies with very high returns generate the majority of the equity risk premium. This observation has important implications for portfolio diversification. Missing out on a handful of highly successful companies can dramatically lower long-term equity portfolio performance. Attempting to pick these big winners *ex ante* is difficult, however, and a strategy of selecting stocks based on skewness has not proven profitable. Research has shown that individual investors tend to hold and have a preference for small, concentrated stock positions and prefer lottery-like, positively skewed stocks.[6] Given the evidence that it is difficult to pick which stocks will generate outsized future returns, this undiversified strategy is a recipe for missing out on the benefits of skewness that accrue to holding a broad index.

Kurtosis: The Weight of the Tails

The fourth moment, kurtosis, measures how thick or heavy the tails of a distribution are. Formally, it is

$$\frac{\mathrm{E}[(X-\mu)^4]}{\sigma^4}. \tag{5.2}$$

For a normal distribution, kurtosis equals 3, so we often use the term *excess kurtosis*, defined as kurtosis minus 3. A high (positive) excess kurtosis means that extreme outcomes, both positive and negative, occur more frequently than under a normal distribution, the distribution has a higher central peak, and—if it is roughly symmetric—there is both greater tail risk and a higher chance of repeated large positive returns.

Both individual stocks and indices show more extreme monthly returns than are consistent with a normal distribution. The degree varies, however: Small-cap stocks tend to have far higher kurtosis than large-cap indices, reflecting their tendency toward large jumps and speculative bursts. For example, Fama and French's market factor shows an excess kurtosis of roughly 7, whereas the small minus big (SMB) and high minus low (HML) factors both have excess kurtosis of around 18.

Comparing Stocks and Indices

Because we already covered the mean and variance of returns in Chapter 4, we now compare the higher moments between different indices and individual-stock-level returns. The contrast between the two types of returns is substantial—at the constituent level, returns are highly nonnormal with high skewness and substantial excess kurtosis. At the index level, aggregation smooths much of this idiosyncratic noise. Firm-specific jumps, crashes, and reversals are muted when many stocks get combined into a single portfolio, leading to much lower index-level skewness. Aggregation, however, does not make index returns normal. Rare systemic shocks

[6] See, for example, Goetzmann and Kumar (2008) and Kumar (2009).

Exhibit 5.5. Skewness and Excess Kurtosis of Monthly Index- and Stock-Level Returns, 1926–2025

	Ibbotson 25	Large Cap	Mid Cap	Small Cap
Skewness (Index)	−0.08	0.08	0.78	1.30
Excess kurtosis (Index)	5.23	7.02	11.77	14.51
Skewness* (Stock)	0.80	1.17	1.64	2.30
Excess kurtosis* (Stock)	7.61	10.96	14.83	23.99

*Stock-level distributions are Winsorized at the 0.01% level for the excess kurtosis results.

Note: The indices are the Ibbotson Indices, and stock-level distributions are from CRSP, using the same definitions as the corresponding Ibbotson Indices.

still affect many stocks at once, so they survive aggregation and continue to generate fat tails at the index level.

Exhibit 5.5 summarizes these relationships using monthly data for US equities from 1926 through 2025. Index-level skewness is close to zero for the portfolios that consist of the largest stocks, with skewness of −0.08 for the Ibbotson 25 and 0.08 for the Large Cap index, but it rises to 0.78 for Mid Cap and 1.30 for Small Cap. At the constituent level, the increase is much sharper, with the stock-level skewness of 0.86 for the Ibbotson 25 stocks, 1.28 for Large Cap stocks, 14.85 for Mid Cap stocks, and 20.50 for Small Cap stocks. The size pattern is very clear in both index and constituent returns, but individual stocks display much more extreme positive asymmetry as expected. Similar patterns persists in the kurtosis while remaining significantly nonnormal across both index and stock returns.

In practical terms, skewness and kurtosis help quantify why markets sometimes move more violently than predicted by the normal distribution. They remind us that diversification reduces but does not fully eliminate tail risk and that real-world investment returns can defy the tidy bell curve we see in many other areas of life. Over the past century, such events as those in October 1929, October 1987, March 2000, and March 2020 have underscored the reality of fat tails and large common shocks in aggregate markets. Conversely, the large gains of a few high-growth firms in each decade are a vivid example of positive skew in individual stock returns.

In summary, mean and variance tell only a part of the story. Skewness reveals the direction of asymmetry in returns, telling us whether it is the extreme gains or losses that dominate. Kurtosis reveals the magnitude of those extremes and how common those extreme events are. Together, they offer a more complete view of return distributions and help us interpret both the calm and the chaos of the past century in markets. Yet it is important to keep in mind that we observe empirical distributions only after the fact, and large events continue to surprise investors. A century-long sample of the US market is enough to demonstrate that standard theoretical models and even relatively sophisticated econometric models do not fully capture the empirical frequency of experiencing rare but large losses and gains. Skewness and kurtosis are essential measures for recalibrating these probabilities using actual data.

Key Takeaways

- US equities have generated large gains in real wealth over the long run, despite occasional long drawdowns.
- Lognormal returns provide a good first-order model for long-run wealth accumulation, but more sophisticated models can account for deviations from lognormality. These are useful for other financial applications, such as option pricing.
- The distribution of individual stock returns and stock index returns is fat tailed compared to the lognormal distribution. This means extreme returns are more likely than the lognormal predicts.
- The distribution of individual stock returns is strongly positively skewed, with a few extreme winners accounting for much of the equity risk premium.
- Broad diversification captures these rare winners without requiring stock selection skill, while concentrated portfolios increase the risk of missing the stocks that matter most.
- Most long-horizon investors will experience at least one major single-day market crash, but these events have not prevented stocks from being a successful long-term investment.
- Understanding tail risk helps investors endure drawdowns and stay invested. It helps investors plan for rare but important shocks.

References

Andersen, Torben G., Nicola Fusari, and Viktor Todorov. 2020. "The Pricing of Tail Risk and the Equity Premium: Evidence from International Option Markets." *Journal of Business & Economic Statistics* 38 (3): 662–78. doi:10.1080/07350015.2018.1564318.

Bachelier, Louis. 1900. "Théorie de la Speculation." *Annales Scientifiques de l'École Normale Supérieure* 17: 21–86. Reprinted in Cootner, Paul H., ed. 1964. *The Random Character of Stock Market Prices*, 17–78. Cambridge, MA: MIT Press.

Bessembinder, Hendrik. 2018. "Do Stocks Outperform Treasury Bills?" *Journal of Financial Economics* 129 (3): 440–57. doi:10.1016/j.jfineco.2018.06.004.

Bessembinder, Hendrik, Te-Feng Chen, Goeun Choi, and K. C. John Wei. 2023. "Long-Term Shareholder Returns: Evidence from 64,000 Global Stocks." *Financial Analysts Journal* 79 (3): 33–63. doi:10.1080/0015198X.2023.2188870.

Bollerslev, Tim, and Viktor Todorov. 2011. "Tails, Fears, and Risk Premia." *Journal of Finance* 66 (6): 2165–211. doi:10.1111/j.1540-6261.2011.01695.x.

Bollerslev, Tim, Viktor Todorov, and Lai Xu. 2015. "Tail Risk Premia and Return Predictability." *Journal of Financial Economics* 118 (1): 113–34. doi:10.1016/j.jfineco.2015.02.010.

Calvet, Laurent E., and Adlai J. Fisher. 1999. "A Multifractal Model of Assets Returns." NYU Working Paper No. FIN-99-072 (November). https://ssrn.com/abstract=1299397.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50 (4): 987–1007.

Fama, Eugene F. 1963. "Mandelbrot and the Stable Paretian Hypothesis." *Journal of Business* 36 (4): 420–29. doi:10.1086/294633.

———. 1965. "The Behavior of Stock-Market Prices." *Journal of Business* 38 (1): 34–105. doi:10.1086/294743.

Gabaix, Xavier. 2012. "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance." *Quarterly Journal of Economics* 127 (2): 645–700. doi:10.1093/qje/qjs001.

Goetzmann, William N., and Alok Kumar. 2008. "Equity Portfolio Diversification." *Review of Finance* 12 (3): 433–63. doi:10.1093/rof/rfn005.

Kaplan, Paul D. 2012. "Déjà Vu All Over Again." In *Frontiers of Modern Asset Allocation*, edited by Paul D. Kaplan, 211–222. Hoboken, NJ: Wiley. doi:10.1002/9781119205401.ch18.

Kendall, Maurice G. 1953. "The Analysis of Economic Time-Series—Part I: Prices." *Journal of the Royal Statistical Society. Series A (General)* 116 (1): 11–25. doi:10.2307/2980947.

Kon, Stanley J. 1984. "Models of Stock Returns—A Comparison." *Journal of Finance* 39 (1): 147–65. doi:10.2307/2327673.

Kumar, Alok. 2009. "Who Gambles in the Stock Market?" *Journal of Finance* 64 (4): 1889–933. doi:10.1111/j.1540-6261.2009.01483.x.

Mandelbrot, Benoit. 1963. "The Variation of Certain Speculative Prices." *Journal of Business* 36 (4): 394–419. doi:10.1086/294632.

McQuarrie, Edward F. 2023. "Stocks for the Long Run? Sometimes Yes, Sometimes No." *Financial Analysts Journal* 80 (1): 12–28. doi:10.1080/0015198X.2023.2268556.

Merton, Robert C. 1976. "Option Pricing When Underlying Stock Returns Are Discontinuous." *Journal of Financial Economics* 3 (1–2): 125–44. doi:10.1016/0304-405X(76)90022-2.

Osborne, M. F. M. 1959. "Brownian Motion in the Stock Market." *Operations Research* 7 (2): 145–73. doi:10.1287/opre.7.2.145.

Santa-Clara, Pedro, and Shu Yan. 2010. "Crashes, Volatility, and the Equity Premium: Lessons from S&P 500 Options." *Review of Economics and Statistics* 92 (2): 435–51. doi:10.1162/rest.2010.11549.

Siegel, Jeremy J. 2022. *Stocks for the Long Run*, 6th ed. New York: McGraw-Hill.

Taleb, Nassim N. 2010. *The Black Swan: The Impact of the Highly Improbable*, 2nd ed. New York: Random House.

———. 2020. *Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications*. New York: STEM Academic Press.

Wachter, Jessica A. 2013. "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?" *Journal of Finance* 68 (3): 987–1035. doi:10.1111/jofi.12018.

CHAPTER 6. BUBBLES, BOOMS, AND CRASHES, 1792–2024

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

James Tyler
Investment Associate, Bridgewater Associates

Introduction

Bubbles are of great interest to investors because they are periods in which stock prices rise dramatically and drop quickly, generating large speculative profits for some and large losses for others. They have spawned many theories, ranging from market constraints to irrational behavior to the fact that bubbles are only identifiable *ex post*: An extreme temporary rise will not be called a bubble unless it is followed by a crash.

A theoretical definition of an asset bubble is a large, positive, temporary deviation from fundamental value. Testing for a bubble under this definition requires a valuation model. Such models must incorporate not only such fundamentals as expected future cash flows but also the complex interaction of differing investor beliefs and behavior.

In this chapter, we settle on a simple, empirical definition that most would agree on. A bubble is a large boom in stock prices followed by a crash. If, for example, a market doubles in value over three years, it would qualify in most people's minds as a stock market boom. If prices then halved in the following years, this would certainly qualify as a crash. Putting these two together defines an empirical bubble. One could choose different thresholds or compounding intervals, but the principle of defining a bubble in terms of price dynamics provides an intuitive framework for enumerating their historical frequencies.

A practical issue for investors is whether a rapid increase in prices predicts a rapid future price drop. The historical rarity of bubbles, at least in the US capital market, makes this question difficult to test empirically. One approach is to expand the sample to the full population of global equity markets using all available time periods, the approach taken in Goetzmann (2015).[1] Another approach is to follow Greenwood, Shleifer, and You (2019) and test whether industry-level price booms are more likely to be followed by crashes. This method is clearly relevant to investors concerned with a boom in a single industry.

[1] See also Goetzmann and Kim (2018), who use a global sample to study negative bubbles.

We begin by using the available history of aggregate US stock market returns to document the historical frequency of booms, crashes, and bubbles. We then use industry portfolios to test for bubble dynamics. We extend the Greenwood et al. (2019) analysis through 2024 and perform an out-of-sample analysis using industry returns from an earlier era in US capital market history, 1871–1938. Finally, we compare the conditional distribution of returns following 12-month industry booms with the entire distribution of returns.

We find that bubbles are extremely rare in both the aggregate index and industry indexes. Negative bubbles, in which a large decline is followed by a large gain, occur slightly more frequently than positive bubbles. Our replication of Greenwood et al. (2019), using earlier data and a sample from 1926 through 2024, closely matches their original findings. Booms do not predict crashes, but crashes are more frequent following booms. Comparing conditional and unconditional distributions shows us why. Booming markets are more likely to be volatile markets in the subsequent period, with a higher likelihood of both large gains and losses.

Booms and Busts in the US Stock Market Since 1792

In this section, we examine the frequency of booms, crashes, and bubbles over the whole course of US stock market history, from 1792 to the present. We use a monthly price index of 100 large-cap stocks, compiled by Finaeon, a data provider specializing in historical financial data. For simplicity, we ignore the substantial return provided by dividends and their re-investment; we are interested primarily in counting big price rises followed by declines.

We define a bubble as a boom followed by a crash that reverses all the market's prior gains. For various boom thresholds—ranging from a 50% gain over three years to a doubling—we count the number of episodes in our overlapping monthly sample and then count how many of those episodes were followed, within one, three, or five years, by either (a) a further boom of the same magnitude or (b) a full reversal that gave back all the earlier gain. We present these counts graphically rather than as tables of conditional probabilities because the raw counts make two facts immediately visible: (1) the rarity of extreme events and (2) the very small number of observations on which any inference about extreme outcomes must rest.

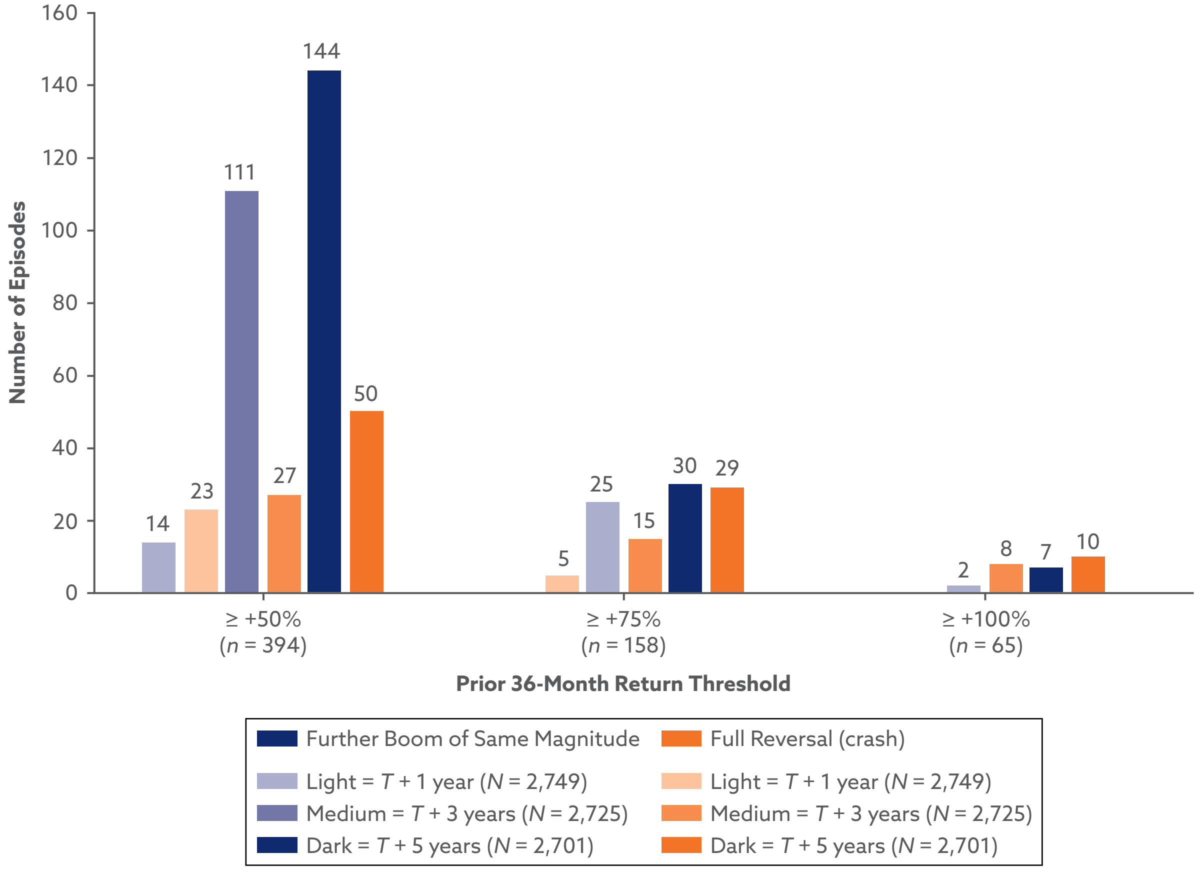
36-Month Booms

Exhibit 6.1 plots episode counts for booms measured over trailing 36-month windows. Orange bars in the figure count the number of bubbles according to the simple definition of a price drop that fully reverses a large prior gain. Consider the 50% threshold: Of the 394 overlapping episodes in which the market rose by at least 50% over three years (an annualized price increase of roughly 14.5%), 144 were followed by another gain of at least 50% within five years (dark blue bar), and only 50 were followed by a full reversal (dark orange bar).

At the three-year horizon, the contrast is even starker: 111 further booms (medium blue bar) versus 27 reversals (medium orange bar). Even within one year, the number of further booms (light blue bar), 14, is comparable to the number of crashes (light orange bar), 23. The visual pattern is clear: Blue bars dominate orange bars at most horizons. For big price runups—75% and 100% over three years—the number of bubbles is a tiny fraction of the sample size, *n*.

At higher thresholds, the same qualitative pattern holds, but the counts become very small. Among the 158 episodes of a 75% three-year gain, 30 were followed by a further 75% gain

Exhibit 6.1. Conditional Outcomes After 36-Month Booms: US Stock Market, 1792–2024



Source: Data from Finaeon, Inc.

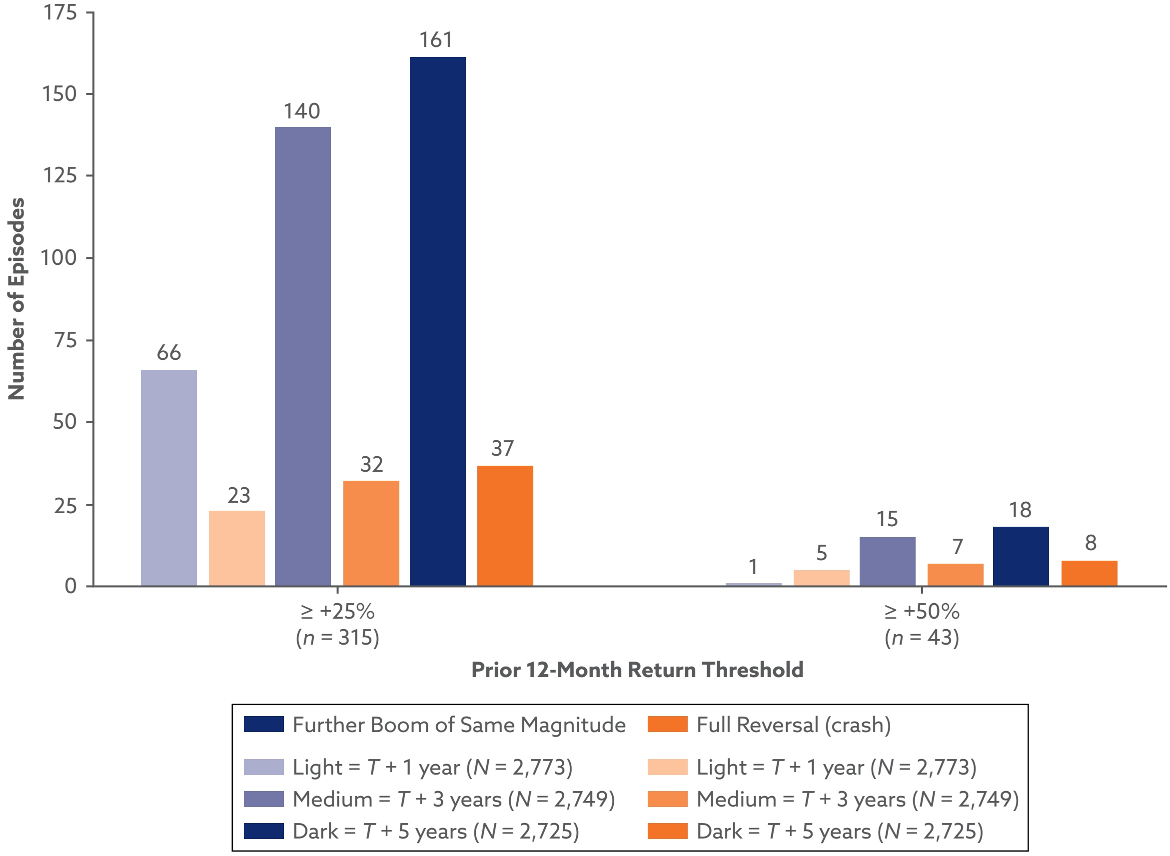
Note: The exhibit shows raw counts of episodes in which a subsequent gain (further boom) or loss (full reversal) of equal magnitude occurred, at forward horizons of one, three, and five years (overlapping monthly observations).

within five years and 29 preceded a full reversal—nearly a coin flip but one based on overlapping observations of a handful of historical events. At the 100% threshold, only 65 episodes occurred in 232 years. Within five years, 7 experienced a further doubling and 10 saw a full reversal. These low counts caution against computing precise odds ratios; the exhibit makes apparent the sparseness of the data.

12-Month Booms

Exhibit 6.2 repeats the exercise for booms measured over 12-month windows. Rapid price increases of this magnitude are rarer still. A gain of 25% or more in a single year occurred 315 times in the sample; 161 of those were followed by another 25% gain within five years, versus 37 full reversals. A gain of 50% or more in 12 months occurred only 43 times. Of those, 18 subsequently gained another 50% within five years, while 8 fully reversed. At higher thresholds, the data are

Exhibit 6.2. Counts of Booms and Crashes After a 12-Month Price Boom: The US Stock Market, 1792–2024



Source: Data from Finaeon, Inc.

Notes: The exhibit shows raw counts, with boom thresholds measured over the preceding 12 months (overlapping monthly observations). The ≥ +75% (*N* = 3) and ≤ +100% (*N* = 0) thresholds are omitted.

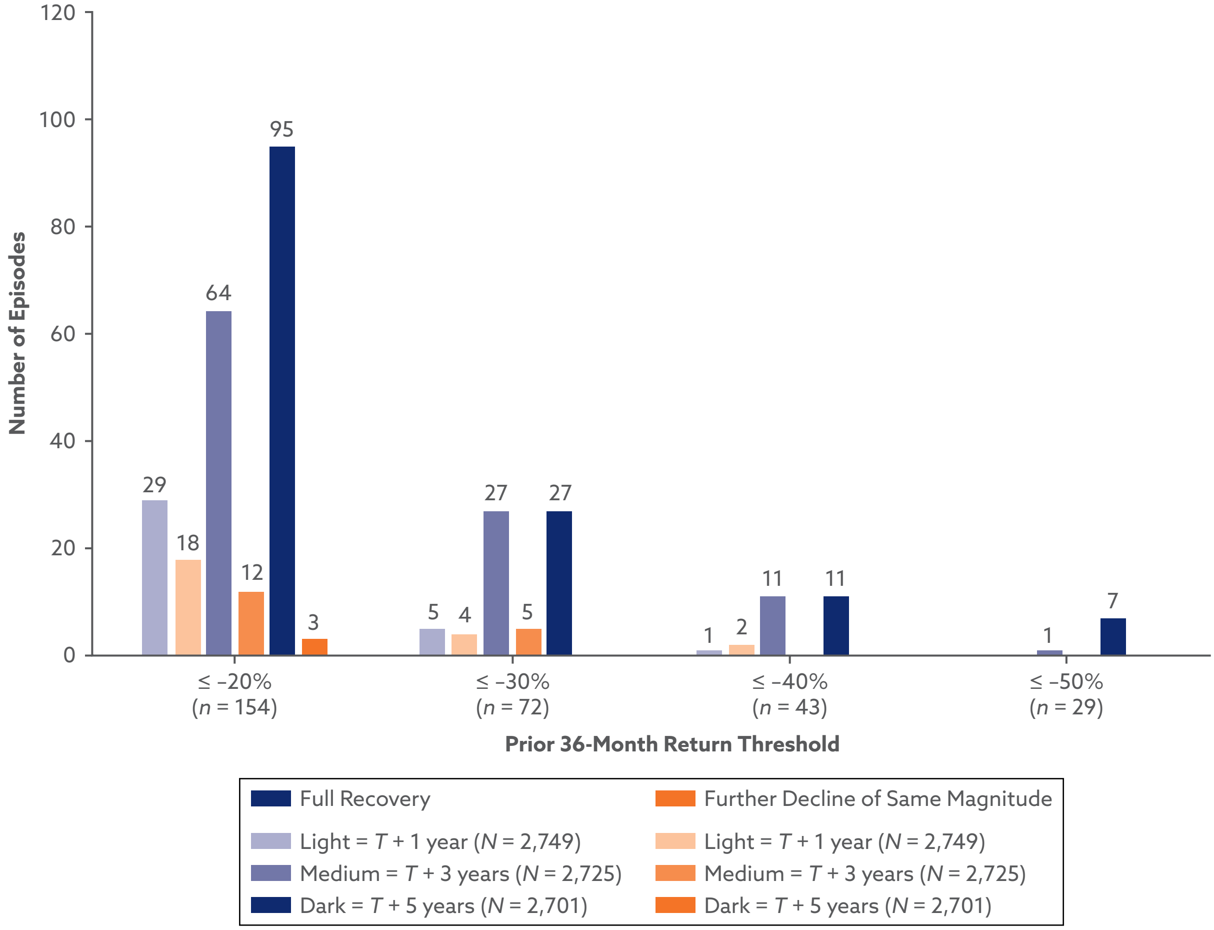
too sparse to report: A 75% 12-month gain occurred just three times, and a doubling in 12 months never occurred. We omit these thresholds from the exhibit. Taken together, Exhibits 6.1 and 6.2 show that booms are more often followed by further booms than by full reversals.

36-Month Crashes

Crashes tell a different story. **Exhibit 6.3** plots episode counts for 36-month declines. Here, the blue bars represent full recoveries—cases in which the market regained all of its prior losses—while the orange bars represent further declines of the same magnitude. The visual asymmetry as the horizon extends is striking: The blue recovery bars grow while the orange further-decline bars shrink toward zero.

After a decline of 20% or more over three years (154 episodes), a full recovery followed within five years 95 times, while only 3 such declines saw a further 20% loss. After a 30% decline

Exhibit 6.3. Counts of Booms and Crashes After a 36-Month Crash: US Stock Market, 1792–2024



Source: Data from Finaeon, Inc.

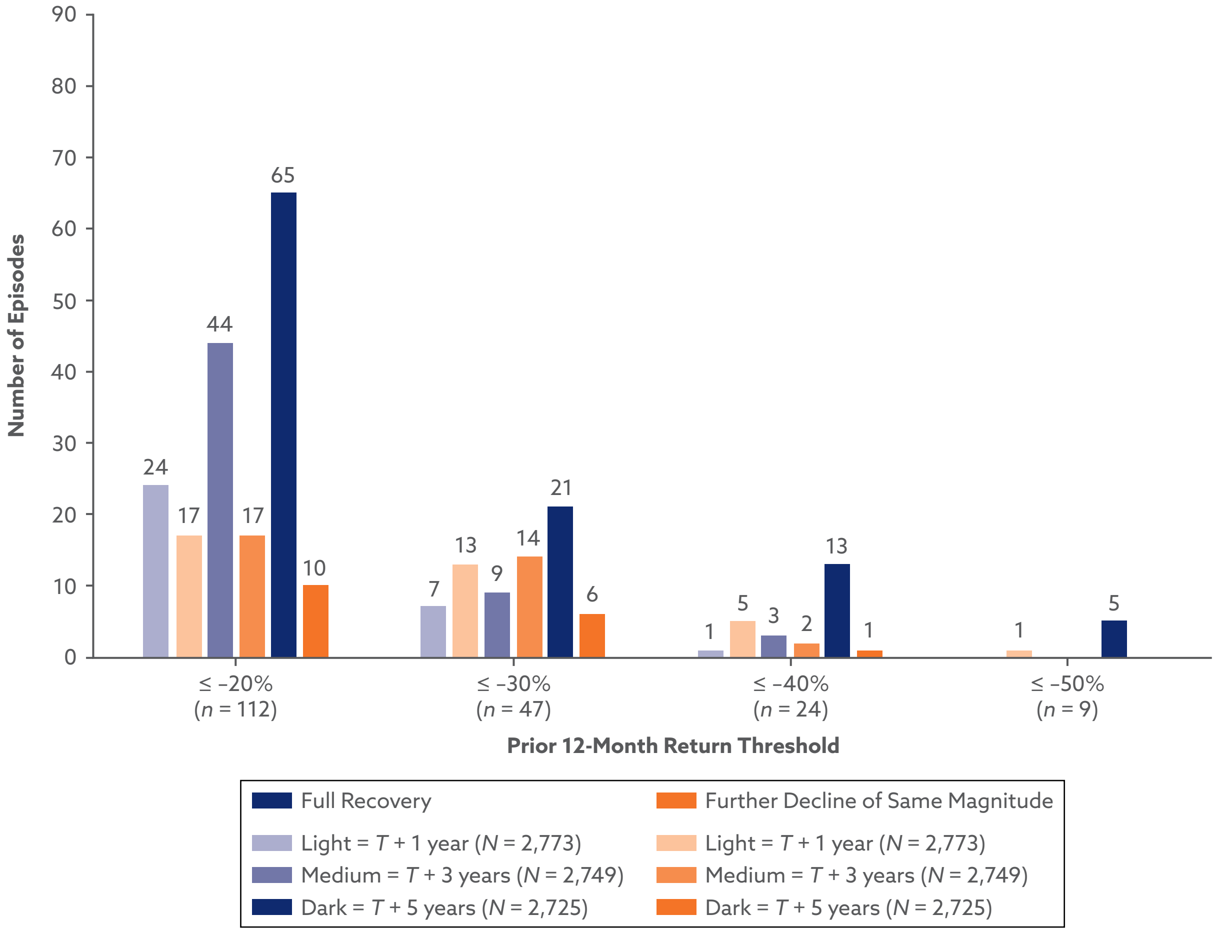
Note: The exhibit shows raw counts of episodes in which a subsequent recovery (reversing the prior decline) or further decline of equal magnitude occurred, at forward horizons of one, three, and five years (overlapping monthly observations).

(72 episodes), 27 recoveries occurred within five years and no further 30% declines occurred. At the 40% and 50% thresholds, the pattern is the same: Recoveries occur and further declines of equal magnitude do not. Not a single episode of a 50% three-year decline was followed by an additional 50% decline at any horizon.

12-Month Crashes

Exhibit 6.4 examines rapid, 12-month crashes—the events most frightening to investors. There were only nine episodes in which the market fell by 50% or more in a single year, less than one-third of 1% of the sample. None of these episodes was followed by a further 50% decline at any horizon, but five of the nine saw full recovery within five years.

Exhibit 6.4. Counts of Booms and Crashes After a 12-Month Crash: US Stock Market, 1792–2024



Source: Data from Finaeon, Inc.

Note: The exhibit shows raw counts, with crash thresholds measured over the preceding 12 months (overlapping monthly observations).

At the 30% threshold (47 episodes), the near-term picture is more sobering: Within one year, 13 episodes saw a further 30% decline versus only 7 recoveries. But within five years, the pattern reverses: 21 full recoveries versus 6 further declines. This finding suggests that rapid crashes carry genuine short-horizon risk of further loss but over longer periods, the dominant outcome is recovery.

Discussion

Several features of these exhibits warrant attention. First, booms are more likely to be followed by further booms than by full reversals. This is true at every threshold and nearly every horizon for the 36-month measurement window and at the 25% and 50% thresholds for the 12-month window. The popular intuition that a large runup in stock prices is a reliable signal of an impending crash is not supported by 232 years of US data.

Second, crashes are even more likely to be followed by recoveries than booms are to be followed by further booms. At the five-year horizon, recoveries dominate further declines at every crash threshold. An investor experiencing an extended decline in the market and considering whether to sell would find little historical support for doing so.

Third, the unconditional probability of a bubble—defined as a boom followed by a full reversal—is very small. At the 50% three-year threshold, there were 50 reversals out of roughly 2,700 overlapping observations, an unconditional frequency below 2%. At the 100% threshold, 10 reversals occurred out of 2,700 observations, roughly one-third of 1%. Reverse bubbles—crashes followed by full recoveries—are somewhat more common, consistent with the general tendency of markets to rise over long horizons.

Fourth, the counts at extreme thresholds are very small, and all are based on overlapping monthly observations. A single historical event such as October 1929 appears in every 36-month window that overlaps with it. The 7 further doublings and 10 reversals at the 100% threshold are not independent observations; they may reflect one or two episodes. We present counts rather than conditional probabilities precisely to make this limitation transparent. Any precise odds ratio computed from a handful of overlapping observations of extreme events should be treated as anecdote, not evidence.

The caveats discussed earlier apply similarly. We have only one historical time series to study, and the fact that it is the US market should make us cautious. Since 1792, the US capital market rose from obscurity to prominence, partly through survival in its early years as an emerging market, recovery after a massive civil war in the nineteenth century, success in two major wars in the twentieth century, and successful management of the global financial crisis of the twenty-first century. Inferences drawn about market recovery from disasters would likely differ if we were using data from pre-revolutionary Russia, Eastern European markets prior to World War II, or markets that survived large shocks but faded in importance with secular changes in the world economic order.

Documenting extreme events in financial history is important for forecasting the distribution of future investment returns. The few large-scale booms and crashes experienced by the US stock market over its history help assess the probability of future tail events but are likely to be conservative estimates of their potential magnitude, if only because they are a very small sample. Bubbles are even more rare than extreme booms and crashes because they require two rare events to occur in succession—although if one causes the other, that changes the probabilities. As we will show next, when we examine industry bubbles, both booms and busts could be correlated because of higher volatility and thus are more likely to occur concurrently.

Industry Bubbles

Thus far we have examined booms, crashes, and bubbles in the historical record of the US market as a whole. Investors are equally interested in the possibility that a single industry is in a bubble. In this section, we extend our boom, crash, and bubble analysis to industry portfolios that make up the US index. Historical data provide a rich cross section of industries extending back to 1871. We use the Cowles (1938) data to test the findings of Greenwood et al. (2019) out of sample, and we use Fama–French industry portfolio data to extend their analysis to the current time. We did not chain the two periods together, because definitions of industries slightly differ. The Cowles data include more industries with presumably finer gradations among them.

As we did earlier with the aggregate market index analysis, we define a boom as a large return over a given time period, either 12 months or 36 months. Because there are many industry indexes, we can construct a large sample of extreme booms and extreme crashes. We thus define a boom as a 100% price increase over a 12-month window, a crash as a 50% price decline, and a bubble as a boom followed by a crash in the subsequent 12 months. To avoid double-counting, if a month is tagged as a boom, the following 12 months cannot initiate a new boom episode. This methodology combines the approaches of Greenwood et al. (2019) and Goetzmann (2015) and provides a simple, transparent way to identify extreme events.

We conducted our analysis using the 49 Fama–French (FF49) total return industry portfolios spanning the period 1926 to 2024 and the 68 Cowles Commission monthly total return industry portfolios, which span the period 1871–1938.[2] The benefit of using these two samples is that by analyzing subsectors of the broad index, we can control for overall market movements, defining a boom as an increase over the market trend as opposed to an absolute return. Greenwood et al. (2019) showed that this control is potentially important because it avoids the results being driven by a handful of market-wide boom and crash periods. Finally, the industry-level analysis addresses an important question for investors: What happens when one industry suddenly booms? Is it likely that a boom in the subsector of the market continues, reverses, or reverts to average industry behavior?

Following Greenwood et al. (2019) with slight modification, a boom is identified when an industry earns both (1) a raw and market-adjusted return of 100% or more over the past two years and (2) a raw return of at least 50% over the past five years. **Exhibit 6.5** reports two event studies conditional on this boom definition.

The booms in both samples had an average growth of 300% over two years. Conditional on this huge gain, however, their average subsequent price path was flat. Panel A of Exhibit 6.5 shows the analysis for the Cowles industries, and Panel B shows the post-1926 period. The graphs look nearly identical. They support the basic finding by Greenwood et al. (2019) that on average, booms are not followed by crashes. The number of booms in each sample was about the same: 52 for the Cowles data, 48 for the Fama–French data.

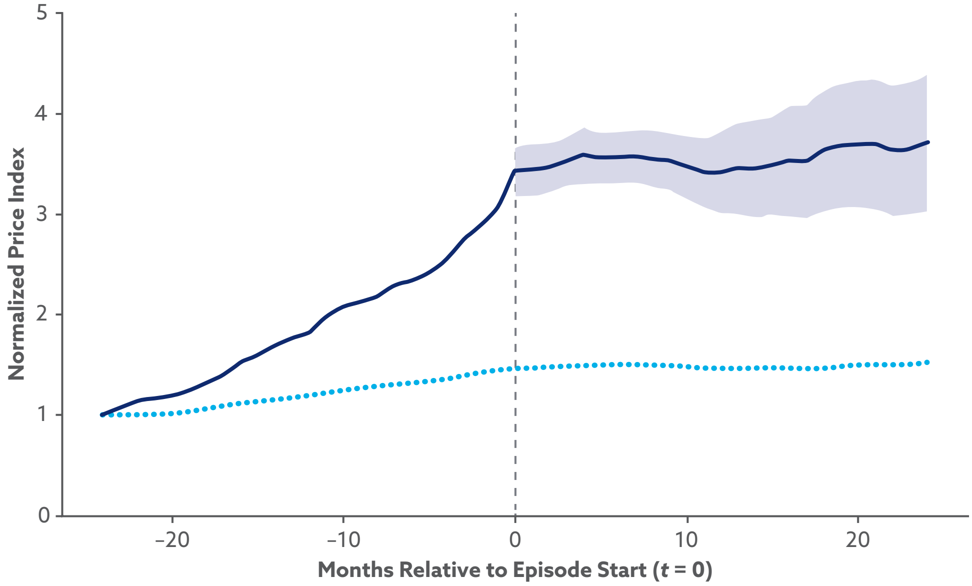
Positive and Negative Bubbles

For another view, we examine both positive and negative bubble episodes. **Exhibit 6.6** plots average paths for booms (12-month doublings) and "negative booms" (12-month halvings) using raw and market-adjusted returns for the Cowles and Fama–French industries. We also use the less stringent boom definition, used in Goetzmann (2015), of an industry doubling or more in 12 months and a crash definition of a decline by half or more. Normalizing the return paths to 1 at the end of their booms or crashes allows us to better see the post-trend cumulative returns. Panels A and C of Exhibit 6.6 show the conditional average path of raw returns

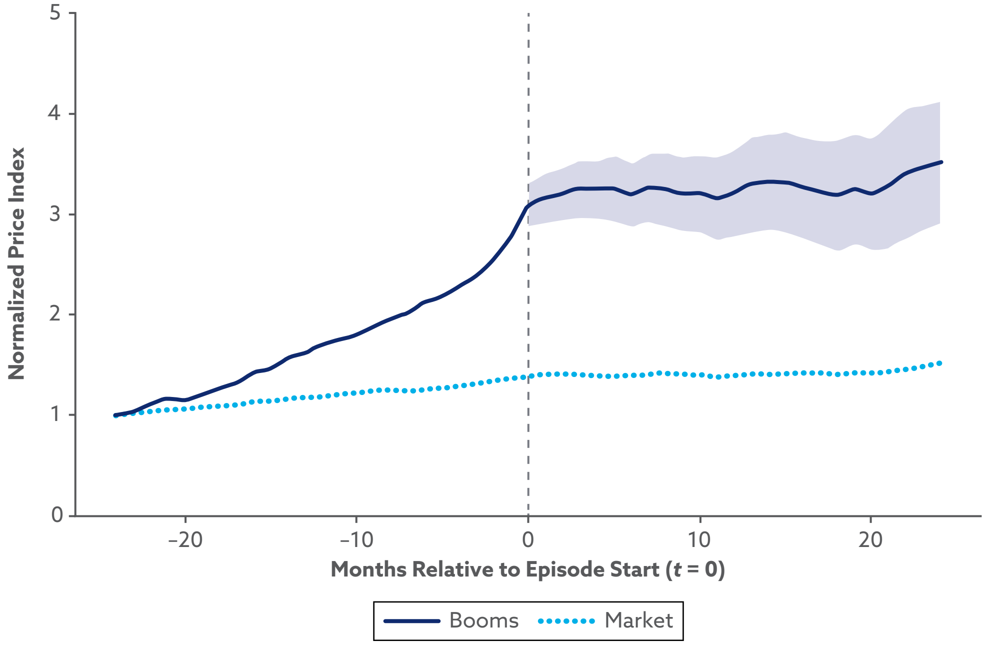
[2] Cowles Commission indices use monthly average prices rather than the typical end-of-month prices.

Exhibit 6.5. Post-Boom Outcomes for Cowles and Fama–French 49 Industries, Average Return for Each Boom Episode from *t* – 12 to *t* + 24

A. Cowles, 1871–1938



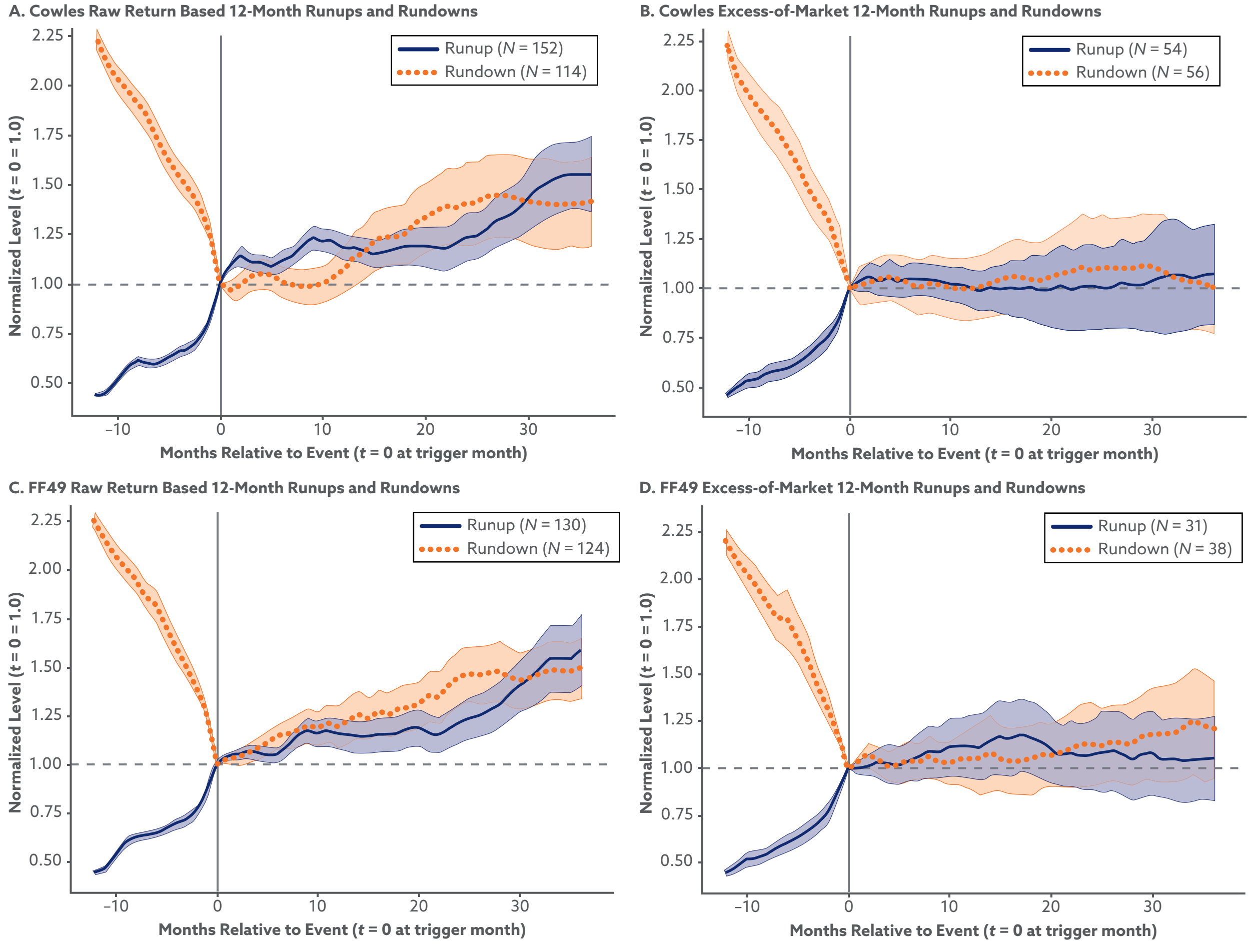
B. FF49, 1926–2024



Sources: Cowles (1938); Kenneth R. French's Data Library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_49_ind_port.html).

Note: The shaded band reports a 95% confidence interval for the cross-episode mean normalized price path.

Exhibit 6.6. Post-Boom and Post-Bust Outcomes Based on Raw and Excess-of-Market Returns, Cowles (1871–1938) and FF49 (1926–2024)



Sources: Cowles (1938); Kenneth R. French's Data Library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_49_ind_port.html).

with confidence bands. Both paths drift up over the subsequent 36 months following their rise or decline, consistent with positive average equity returns for the US market over the sample periods. Confidence bands indicate that for most horizons, the null of post-event equality of boom paths and crash paths cannot be rejected, except for the first few months following the event date.

Panels B and D of Exhibit 6.6 define booms and crashes as, respectively, a doubling or halving in excess returns over the market. This definition significantly reduces the sample size; however, it removes the market trend from the conditional return paths. For both the early sample and the more recent sample, the average paths following booms and crashes are indistinguishable from the market. Notice a subtle difference between the raw and excess figures: In the Cowles sample (Panels A and B), the raw return paths exhibit a lot of time-series variation even though the sample size is larger than the excess return sample. This variation is likely caused by a handful of major market events that made a number of industries boom or crash at the same time. Defining booms and crashes in terms of excess returns over the market mitigates some of this sample selection bias.

A possible survivor bias is introduced by dropping industries when they disappear from the sample, resulting in averaging only across surviving industries in event time. To address this issue, we made the conservative assumption that capital in a discontinued industry was moved to the market portfolio at the end of its last monthly industry observation. The results are little changed by this assumption.

The terminal date of 1938 also censors industries from the event study. To test the effect of this, we analyzed a subset of booms and crashes prior to 1936, allowing us a full three-year potential history. The results do not differ significantly from those reported. To further investigate the source of the slightly positive paths in the excess return figures, we plotted the median of the cumulative excess return indices in each period. Doing so resulted in a terminal median value close to zero.

The results are strikingly similar across the two datasets: On average, prices rise into the boom and then on average level off, with mean paths roughly in line with the market thereafter. In the Cowles data, prices typically go flat after a sharp increase. In the more modern Fama–French data, the price evolution is closer to a smooth transition from the initial runup.

Conditional Distributions

In the analysis so far, we have studied boom-conditional and crash-conditional return distributions. In this section, we explicitly test whether crashes are more likely if there is a preceding boom. We do this by comparing conditional distributions with unconditional distributions—that is, with the entire set of historical industry return paths, not only those that condition on a previous boom or crash. We test the null hypothesis that post-boom return paths are not unusual when compared with the large, unconditional sample.

Greenwood et al. (2019) noted that even though booms do not on average predict crashes, they are associated with a higher subsequent probability of a crash. This apparent contradiction results from the fact that by picking industries that had dramatic runups and declines, we are conditioning on high volatility. In this section, we investigate the shape of the conditional distributions compared to the unconditional distribution of industry returns and test for conditional volatility and skewness using the unconditional distribution as a benchmark.

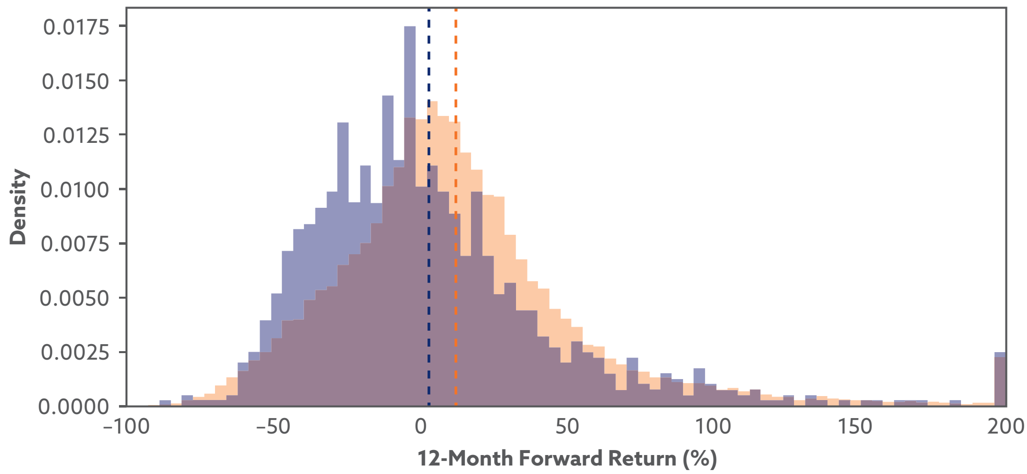
Exhibit 6.7 plots the distribution of industry returns conditional on a prior 12-month boom in blue and the entire distribution of unconditional 12-month returns in orange. As previously noted, the sample of unconditional returns includes multiple overlapping periods; conditional 12-month returns do not overlap.

Panels A and B of Exhibit 6.7 use the Cowles sample, which contains 68 industries, and Panels C and D use the Fama–French sample of 49 industries. Panels A and C show results for raw returns; Panels B and D show results for excess returns. The dashed vertical lines mark the means of each distribution.

Notice that the distribution of conditional raw returns for the 1926–2024 period (Panel C) is wider and shifted to the left compared with the unconditional distribution. The difference in means is large and significant, and the difference in modes in Panel C is also dramatic.

Exhibit 6.7. Post-Boom Outcomes Based on Raw and Excess-of-Market Returns, Clipped to (–100%, +200%): Cowles (1871–1938) and Fama–French 49 (1926–2024)

A. Cowles Raw Returns



B. Cowles Excess-of-Market Returns

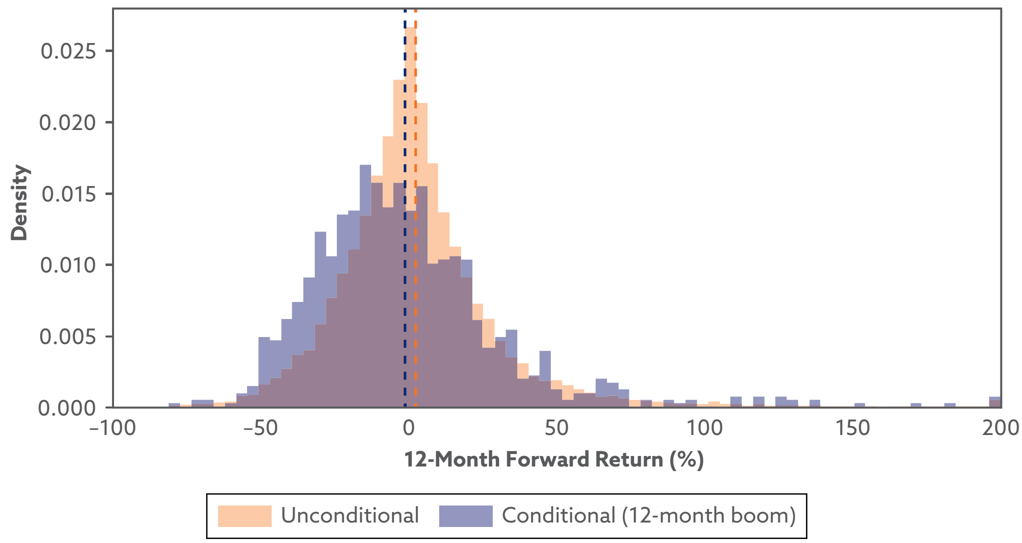
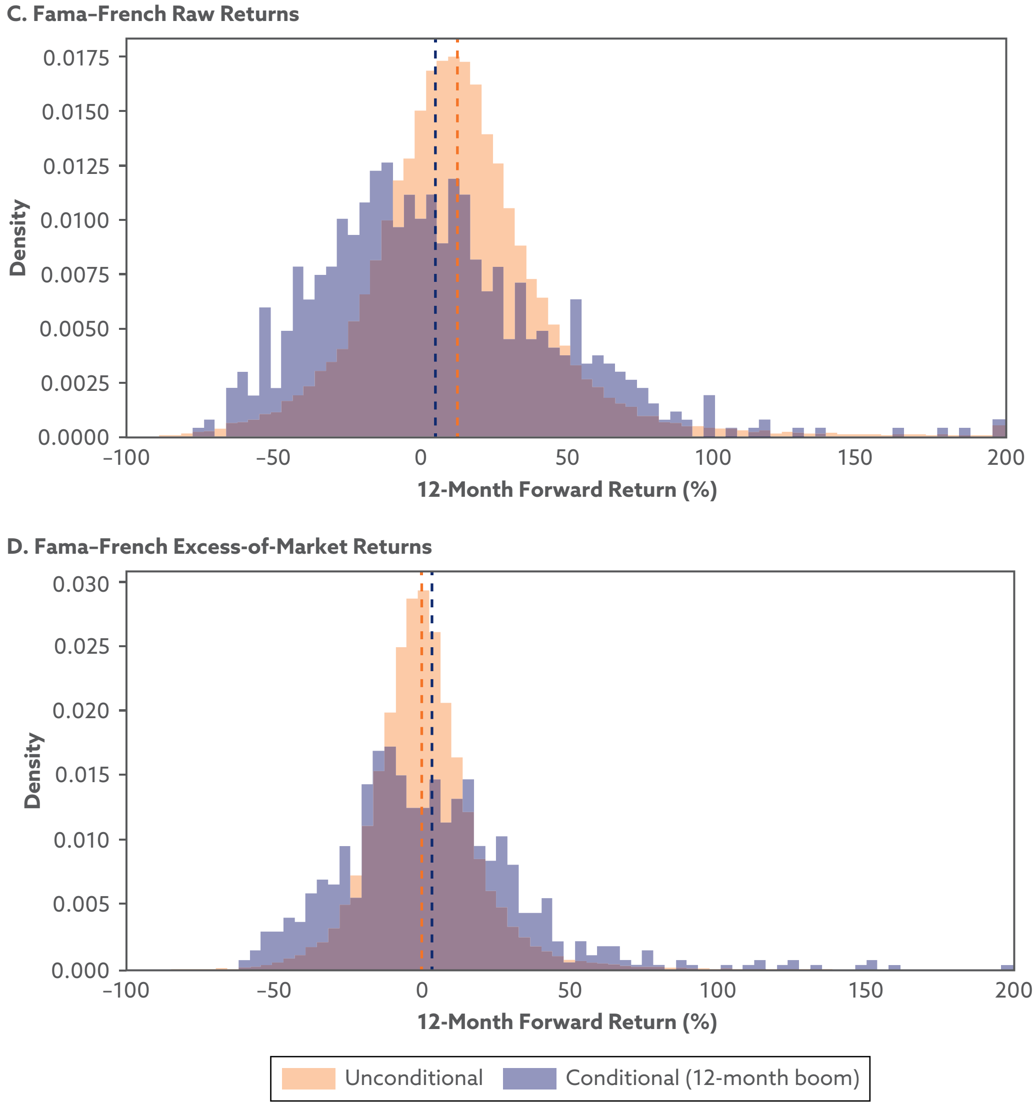


Exhibit 6.7. Post-Boom Outcomes Based on Raw and Excess-of-Market Returns, Clipped to (–100%, +200%): Cowles (1871–1938) and Fama–French 49 (1926–2024) (*continued*)



Sources: Cowles (1938); Kenneth R. French's Data Library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_49_ind_port.html).

Panel C suggests that at the industry level, 12-month booms are a strong predictor of low excess returns in the following 12 months. The distribution of subsequent excess returns shown in Panel D, however, suggests this is not necessarily the case. The mean of the subsequent excess returns is slightly higher than the unconditional mean. The wide spread remains:

The probability of a crash after an industry boom is much higher than the unconditional probability, but so is the probability of another boom.

Conditioning on a dramatic boom selects for either high-volatility industries or industries that are in a transitory but persistent phase of high volatility (see **Exhibit 6.8**).

Exhibit 6.8. Post-Bust Outcomes Based on Raw and Excess-of-Market Returns, Clipped to (–100%, +200%): Cowles (1871–1938) and Fama–French 49 Industries (1926–2024)

A. Cowles Raw Returns

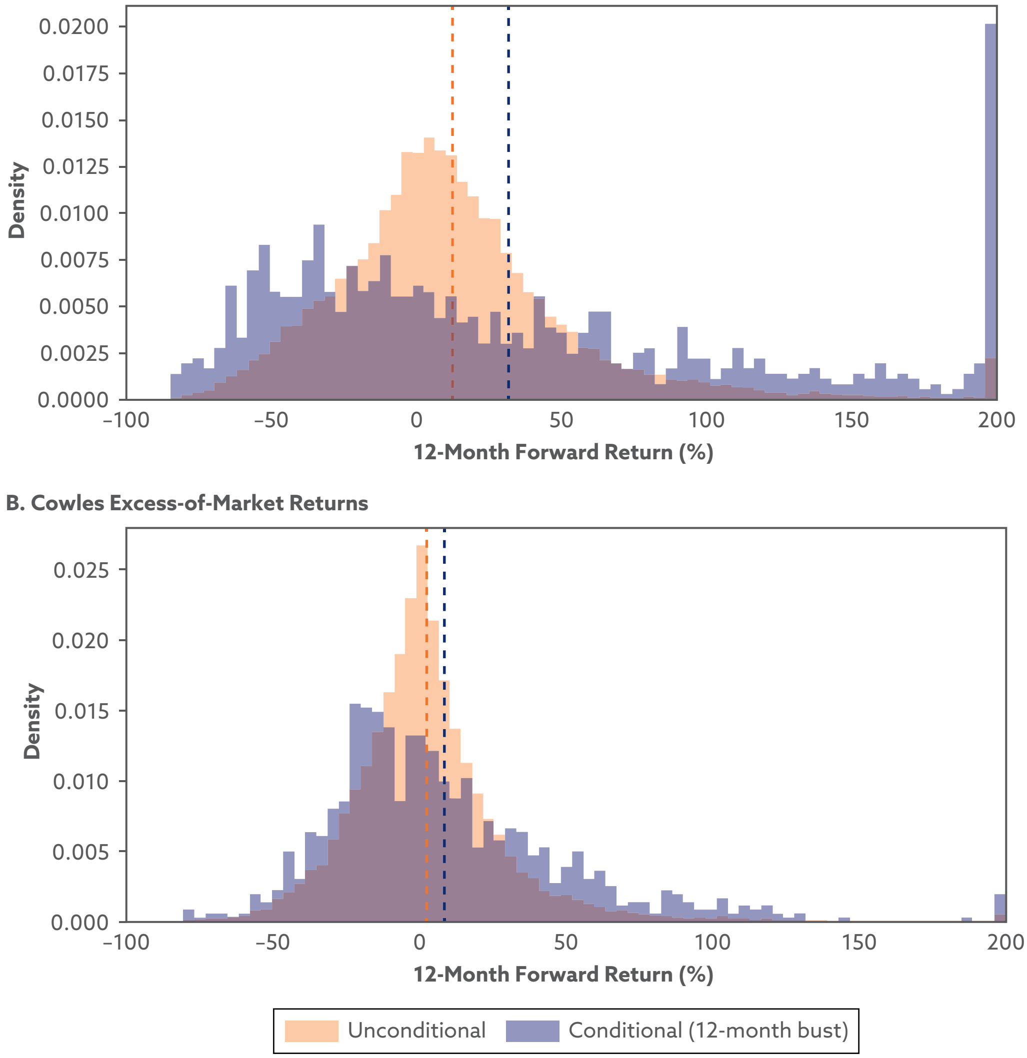
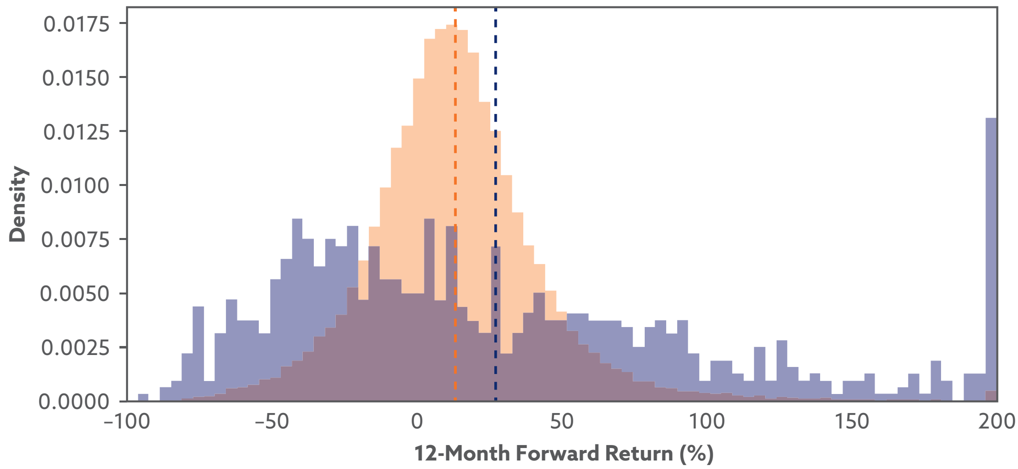
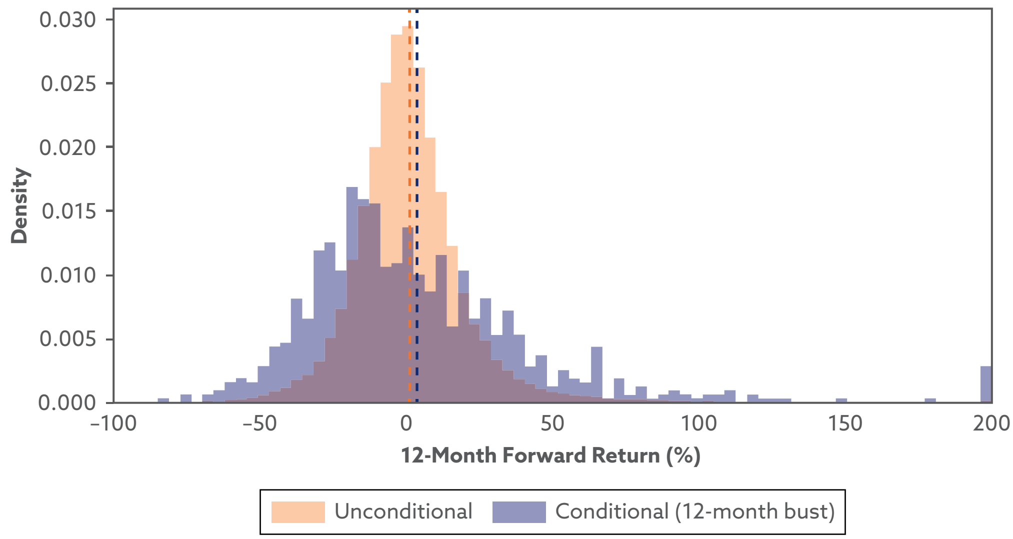


Exhibit 6.8. Post-Bust Outcomes Based on Raw and Excess-of-Market Returns, Clipped to (–100%, +200%): Cowles (1871–1938) and Fama–French 49 Industries (1926–2024) (*continued*)

C. Fama–French Raw Returns



D. Fama–French Excess-of-Market Returns



Sources: Cowles (1938); Kenneth R. French's Data Library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_49_ind_port.html).

Conclusion

This chapter uses an empirical approach to define stock market booms, crashes, and bubbles. Large-scale booms and crashes have been relatively rare in US market history. Bubbles, defined as a boom followed by a crash, are by definition even rarer—one rare event must occur right after another rare event. The probability of this happening is heightened in periods of higher volatility, but so is the opposite sequence—a negative bubble in which a crash is followed by a boom. Most important for investors is that the bubbles and crashes the US market experienced over more than two centuries of data have not been fatal to either the existence of the market or the fortunes of long-term investors.

There are many potential explanations for this continuity. Our analysis of the aggregate US index studies one market that grew to global dominance by the end of the twentieth century. Indeed, it is the first market for which high-quality, long-term equity return data were collected and studied by financial economists (e.g., Cowles 1938). Scholars studying the cross-sectional history of global equity markets (e.g., Jorion and Goetzmann 1999; Dimson, Marsh, and Staunton 2002; Siegel 1994) have had to deal with breaks caused by war, political disruption, and rejection of the market economy. Ultimately, analysis of international equity data showed that the United States was an unusually good performer but that the equity premium is widespread and robust, even taking into account the major shocks and disruptions experienced over the past two centuries.

The problem with analyzing a single market is the rarity of extreme events, which makes it difficult to test predictive models. In practical terms, the conclusion we can draw about booms and crashes based on a single, albeit multicentury return series is that they are infrequent and hard to predict with prior price trends.

The industry-level analysis in this chapter provides a richer test bed for boom and crash prediction and the study of bubbles. The rise and fall of industrial sectors are a fundamental prediction of Schumpeter (1942). For much of its history, the US stock market has channeled investment into innovative technologies that rise, succeed, and decline over time, replaced by successive waves of competition. Pástor and Veronesi (2009) model this Schumpeterian process and find that technologies characterized by high uncertainty and fast adoption are prone to bubbles.

The industry-level analysis shows that booming (and crashing) industries are associated with subsequent high volatility, making both crashes and repeat booms more likely. Once one controls for market trends, however, post-boom performance does not differ significantly from unconditional returns. For crashing industries, rebounds are slightly more likely than further declines. This information is potentially useful for investors considering tactical industry allocations in real time.

Key Takeaways

- True bubbles are rare. A bubble, defined as a large boom fully reversed by a crash, requires two rare events in sequence. In more than two centuries of US data, such full boom-and-bust episodes represent a tiny fraction of all return observations.
- Even very large multiyear or one-year runups in the US equity market have historically been more likely to be followed by further gains than by full reversals.

- Crashes (whether or not preceded by big booms) are rare, and deep crashes are usually followed by recovery rather than further collapse.
- At the industry level, dramatic runups are followed by a much wider distribution of subsequent returns: Both crashes and repeat booms become more likely. Thus, industry booms signal higher volatility (up or down), not necessarily an impending crash.
- The US market is exceptional in that it has survived wars, crises, and regime changes. The fact of its survival means that crash and recovery statistics from this single successful market could understate the range of outcomes in less fortunate markets. It is difficult to infer precise probabilities from only one run of history in the most successful market in the world.

References

Cowles, Alfred. 1938. *Common Stock Indexes, 1871–1937*. Bloomington, IN: Principia Press.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2002. "Long-Run Global Capital Market Returns and Risk Premia." Working paper (February). doi:10.2139/ssrn.299335.

Goetzmann, William N. 2015. "Bubble Investing: Learning from History." NBER Working Paper 21693 (October). www.nber.org/papers/w21693.

Goetzmann, William N., and Dasol Kim. 2018. "Negative Bubbles: What Happens After a Crash." *European Financial Management* 24 (2): 171–91. doi:10.1111/eufm.12164.

Greenwood, Robin, Andrei Shleifer, and Yang You. 2019. "Bubbles for Fama." *Journal of Financial Economics* 131 (1): 20–43. doi:10.1016/j.jfineco.2018.09.002.

Jorion, Philippe, and William N. Goetzmann. 1999. "Global Stock Markets in the Twentieth Century." *Journal of Finance* 54 (3): 953–80. doi:10.1111/0022-1082.00133.

Pástor, Ľuboš, and Pietro Veronesi. 2009. "Technological Revolutions and Stock Prices." *American Economic Review* 99 (4): 1451–83. doi:10.1257/aer.99.4.1451.

Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy*. London: Routledge.

Siegel, Jeremy J. 1994. *Stocks for the Long Run: A Guide to Selecting Markets for Long-Term Growth*, 1st ed. Irwin Professional Publishing.

CHAPTER 7. US TREASURY YIELDS AND RETURNS

Thomas S. Coleman, PhD
Senior Lecturer and Executive Director, Center for Economic Policy, Harris School of Public Policy, University of Chicago

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

US Treasury Yields: What They Are and Why We Care

US Treasury bond yields are the foundation of finance—the "riskless" (no risk of default) yields across all time horizons from one week to 30 years. The Treasury market makes up the largest category of bonds in the United States: 61% of the $48.9 trillion fixed-income market in 2025. All other such markets (corporate bonds, municipal bonds, asset-backed securities, home mortgage rates) are priced *relative* to Treasury yields.

The yield-to-maturity for a bond is the discount rate or internal rate of return that balances the price we pay today (the present value) against future coupon cash flows and the return of principal at maturity. Yields allow us to compare bonds with different prices and future coupons. In a real sense, the yield-to-maturity measures the "price" of a bond, with a higher yield representing a lower price today—a lower dollar price relative to higher future coupons.

When the Treasury issues new bonds, it sets the coupon so that the price of the bond is at or close to $100; the coupon is set higher or lower so that investors are just willing to pay $100 in return for those future coupons. A bond with a price of $100 is called a *par bond* because it is trading at par—$100 per $100 of principal to be returned at maturity. In a period of high market yields (as in the early 1980s in the United States), the Treasury will need to promise a high coupon, and in a period of low yields (as in the 2010s), the coupons will be low. Par bonds and the yield or coupon on such par bonds are the fundamental measure of the Treasury market, the reference yield from which all other securities are priced.

Par bonds and thus par bond yields often are not available directly from the market for all maturities on all dates. As bonds age and market yields go up or down, the price of a bond (originally at or near $100) will go down (as yields go up) or up (as yields go down). For any particular day, a variety of bonds issued in the past will have prices quite different from $100. In other words, although investors always want to know the par bond yield, there are almost never exact par bonds for the maturities we want.

The goal of our Treasury yield curve project is to estimate synthetic par bond yields from the prices and yields of traded bonds for each month—to estimate a full yield or forward curve from which we can then calculate par bond yields for any desired maturity. The Center for Research in Security Prices (CRSP), originally founded at the University of Chicago and now part of Morningstar, has collected prices on all outstanding Treasury bonds for each month from December 1925 to the present. We use these traded bond prices to estimate a full yield or forward curve, as outlined in the next section. From the estimated forward curve, we can calculate par bond yields.

Price, Yield, and Par Bonds

Investors buy and sell bonds based on their price; that is a truism. But we cannot use raw prices to compare one bond with another because bonds generally involve different-size cash flows that occur at different times. For example, is it better to pay $101.93 today for a 7% bond ($7 in the next year plus $100 at the end of the year) or a lower upfront price of $100 for a 5% bond with lower future cash flows? We need to evaluate the two bonds using a common unit of measurement, determining whether the higher upfront price of the 7% bond is justified by the larger future cash flows.

It turns out that these two bonds provide (for all intents and purposes) the same value, but we cannot answer the question without the *yield-to-maturity* (or yield). This calculation provides the common unit of measurement for balancing upfront price against the size and timing of future cash flows. We can think of the yield as the "promised or expected earnings per dollar invested"—in other words, what you will earn if you hold the investment until maturity.

For bonds that pay coupons twice a year (such as Treasury bonds and notes),[1] the formula is

$$\text{Price} = \frac{c/2}{1+y/2} + \frac{c/2}{(1+y/2)^2} + \cdots + \frac{c/2}{(1+y/2)^{2n}} + \frac{100}{(1+y/2)^{2n}}.$$

In our example, both investments (wherein you pay $101.93 for the 7% bond or $100 for the 5% bond) yield 5%. They are in an important sense the same "price" because they yield the same return. Note that because the yield, *y*, appears in the denominator, *as yields go up, prices go down*.

For displaying our estimated yield curves, we want to graph the yield on the hypothetical or synthetic bond versus its maturity. The most convenient method of ascertaining that yield is to calculate the yield on a coupon bond that trades at $100, where the coupon is adjusted until it and the yield are the same. These calculations, called *par bond yields*, are the basic building block we use in most of the exhibits in this chapter.

---

[1] Treasuries pay coupons semiannually, and the market convention is to calculate semiannually compounded yields for bonds with one or more coupon periods. For bills and bonds with six months or less until maturity, markets use the simple yield annualized at a rate of actual days divided by 360. In this study, for consistency across the curve, we calculate all our yields on a semiannually compounded basis.

Estimation of Bond Yields and Returns

Our goals for US Treasury markets are straightforward: Estimate returns for selected one-bond "portfolios" with maturities of, say, three months, 2 years, 10 years, and 20 years.[2] The challenges are manifold. We wish to buy a bond today (say, the bond with two years to maturity) and sell it in one month (when it has one year and 11 months to maturity). This procedure requires an estimated forward-rate or yield curve to create synthetic exact-maturity bonds with at-market coupons and then to revalue those bonds one month later. Estimating such a forward curve for each month over the past 100 years involves the challenge of choosing a flexible forward curve methodology that works for both the many bonds of the modern era (more than 400 now) and the small number of bonds in the 1920s. We choose a piecewise twisted forward curve, discussed in more detail in Chapter 8.

For our forward curve methodology, we build on and extend the methodology used in Coleman, Fisher, and Ibbotson (1989, 1992, 1993). In those studies, we assumed forward rates were constant (flat forwards) over preset maturity periods. In this chapter, we allow forward rates to twist and we more flexibly determine the maturity or knot points.

Government bonds today are simple: Bills pay $100 upon maturity (and are issued with maturities of one year or less), and notes and bonds pay coupons semiannually plus $100 on the stated maturity date. Historically, the bond market has had many quirks, which we now discuss.

Tax Regimes and Tax Status Changes from the 1920s to the 1940s

In the first part of the twentieth century the Treasury issued bonds with varying tax treatment: partially tax-exempt bonds in the 1920s, fully tax-exempt bonds in the 1930s, and taxable bonds starting in the 1940s. Out of necessity, we estimate three different curves over different periods: partially exempt from 1926 to 1931, fully exempt from 1932 to mid-1941, and taxable from mid-1941 to present. In Chapter 8, we detail how we stitch together the disparate regimes to provide the most consistent set of returns and curves available.

Callable Bonds and Noncallable Yields Through the Early 1990s

Through 1984 the Treasury issued mostly callable long-dated bonds—callable at the option of the Treasury, at $100. For example, the 11.75% coupon bonds maturing on 15 November 2014 were callable starting in November 2009; a five-year call period was common, but the period could be longer or shorter. Callable bonds were outstanding through the early 2000s. Traditionally, an ad hoc "yield-to-worst-call" rule was used for calculating an ad hoc (but incorrect) yield for the bond. This rule treats the next call date as if it were the maturity date if the bond is trading at a premium today. If the bond is trading at a discount, the rule treats the final maturity date as the maturity date.

[2] Our detailed exhibits include a wider range of maturities: one, three, and six months; 1, 2, 3, 5, 7, 10, and 20 years; and the longest bond.

Instead of "yield-to-worst," we estimate the value of a European call to the first (or next) call date, a far better approximation for the callable nature of such bonds. This approximation allows us to strip out the value of the call feature and estimate the pure discount forward curve. Thus, our forward and par yield curves are estimates of discount rates for fixed cash flows after removing the call option value.

Flower Bonds Through the 1970s

Most long-dated bonds issued before 1973 could be redeemed at par for payment of estate taxes, giving them the name "flower bonds." Starting in 1973, the Treasury issued longer-dated non-flower bonds. Prior to mid-1973, we estimate a curve including flower bonds, giving a "flower-bond long end" for this period. In mid-1973, we switch to non-flower bonds and exclude flower bonds. In Chapter 8, we detail how we stitch together the curves to produce the most consistent curve possible.

Returns with Our Estimated Forward Curves

We construct par bond yield curves, which we display in the next section. We can also, of course, calculate prices of bonds. This approach enables us to calculate prices and returns for the following:

- This month, buy a new par bond for $100.
- Next month, sell the bond, which is one month shorter.

The trick here is that by using our estimated forward curves to calculate our bonds, we can start with a bond this month that is exactly $100 with, say, a maturity of exactly one year. Next month, the bond's maturity will be only 11 months, but we can use next month's curve to estimate what the price of that old bond would be. We can decompose the *total* return into an *income return* component equal to the promised yield (par bond yield) and the *return in excess of promised yield*. The following sidebar gives more detail.

Calculating Returns

We wish to calculate returns for a variety of bonds (a three-month bond, a one-year bond, and so on), and we wish to decompose the *total return* into an *income return* and a *return in excess of yield*. Because no coupons are paid in the one-month holding period for which we calculate the month-to-month returns, the total return is simply the change in the price of the bond: For a two-year bond from January to February, we need to buy a two-year par bond (trading at $100) at the end of January and sell the same bond at the end of February. The result is now a one-year and 11-month bond. We price the original bond (calculate the coupon or par bond yield, Y_{Jan}^{pb}) from our January curve, and then we price the one-year and 11-month bond (calculate the new price, P_{Feb}, for the January bond we started with) from our February curve.

The Coleman Fisher Ibbotson (henceforth CFI) total return (TR) is simple:

$$\mathrm{TR}_{CFI} = \frac{P_{Feb}}{100} - 1.$$

The income return (Inc) is the par bond yield from January, Y_{Jan}^{pb}, converted to a monthly return:

$$\mathrm{Inc}_{CFI} = (1 + Y_{Jan}^{pb}/200)^{AD/182.625} - 1,$$

where *AD* is the actual number of days from end-January to end-February. The return in excess of yield, RxY_{CFI}, is the total return minus the income return—the "minus" meaning "geometric subtraction" of dividing 1 plus total return by 1 plus income return and then subtracting 1:

$$RxY_{CFI} = \frac{1 + \mathrm{TR}}{1 + \mathrm{Inc}} - 1.$$

US Treasury Bond Yields and Returns, 1926–2025

In addition to presenting our analysis for the full century-long period of 1926–2025, we divide it into two roughly equal subperiods (1925–1980 and 1981–2025) because the subperiods differed radically in the economic environment for bonds and bills.

The Full Period, 1926–2025

From our estimated forward curves, we derive a variety of yields and rates:

- CFI US Treasury bill, note, and bond yields: par bond yields, the yield on a bond that is trading at $100 (with yield = coupon)
- CFI US Treasury bill, note, and bond total returns: the return from buying a bill or bond at Month 0 and then selling the same bond one month later
- CFI income return: the yield-to-maturity or promised yield of the par bond
- CFI return in excess of yield: the total return minus the yield-to-maturity (promised yield) of the bond when purchased (with "minus" meaning the geometric subtraction as detailed previously)

Exhibit 7.1 shows the annualized average of monthly values (and, where appropriate, the annualized standard deviation) over the full century from 1926 to 2025 for these series across the yield curve. The entry for "Long" is the yield for the longest bond for each month, which varied from a low of 14.5 years (in 1980) to 40 years (in 1955).

Exhibit 7.1. Par Bond Yields and Returns, 1926–2025

	1 M	3 M	6 M	1 Yr	2 Yr	3 Yr	5 Yr	7 Yr	10 Yr	20 Yr	Long
Par bond yield (average)	3.25	3.41	3.54	3.72	3.95	4.12	4.35	4.51	4.64	4.86	4.86
Compound Annual Return											
Total return	3.26	3.52	3.76	4.03	4.37	4.56	4.84	4.89	4.97	4.88	4.78
Income return	3.26	3.41	3.54	3.73	3.96	4.14	4.37	4.54	4.67	4.90	4.90
Return in excess of yield	0.00	0.10	0.21	0.29	0.39	0.41	0.45	0.34	0.29	−0.01	−0.12
Arithmetic Mean Annual Return											
Total return	3.30	3.57	3.81	4.09	4.44	4.66	5.00	5.10	5.27	5.40	5.51
Annualized Standard Deviation											
Total return	0.87	1.00	1.14	1.53	2.41	3.24	4.53	5.52	6.68	9.12	10.32
Income return	0.87	0.93	0.93	0.92	0.90	0.88	0.85	0.84	0.82	0.79	0.78
Return in excess of yield	0.00	0.26	0.60	1.19	2.19	3.07	4.38	5.39	6.56	9.00	10.21

Source: Author calculations from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in Chapter 8.

Note that for the *average* total return, most of the return is generated by the *yield* return, with the return in excess of yield being not far from zero (because yields at the end of the century were similar to those at the beginning). For volatility, the situation is reversed, with almost all the volatility resulting from variation in the return in excess of yield. One can think of this in terms of Aesop's parable of the hare and the tortoise: Returns in excess of yield may sprint ahead in any given month, but the tortoise of steady accumulation of income returns pays off in the long run.

It is important to understand the relation between changes in yields and returns. As yields go up, prices go down. One can understand this dynamic either intuitively—by recognizing (as pointed out earlier) that a high yield corresponds to a low price, so a rising yield means a falling price—or by the present value (PV) formula discussed earlier. It is also crucial to understand that when yields go up and prices down, prices for longer bonds go down by more: Longer bonds are more sensitive to yields than shorter bonds. This interest rate elasticity (technically a semi-elasticity, also called the *modified duration*) provides the numerical link between changes in a bond's yield and price. We will discuss this relationship in more detail later. The important fact is that

$$\%\Delta \text{ in Price} \approx \text{ModDur} \times (-\Delta\text{Yield}). \tag{7.1}$$

With a modified duration of 4.9 (roughly the value for a five-year zero-coupon bond at a 5% yield, 4.9 ≈ 5/1.025), a change in yield from 5% to 5.10% (an increase of 0.10 percentage points) would mean the bond price falls by roughly 0.49%:

$$-0.49\% = 4.9 \times (-0.10). \tag{7.2}$$

The modified duration for a one-year bond is slightly less than 1.0, so the same rise in yield would lead to a roughly 0.1% price decrease for a one-year bond.

We can (conceptually) decompose the total return for a bond that we buy in January and sell in February into two parts: the income or accrual from holding the bond for a month (the *income return* discussed earlier) and the price change resulting from changing markets or yields (the *return in excess of yield*). The income return is approximately the yield divided by 12 (the exact formula was given previously), and the excess return is approximately −ModDur × ΔYield,[3] which gives

$$\text{Total return} \approx \frac{\text{Yield}}{12} + \text{ModDur} \times (-\Delta\text{Yield}). \tag{7.3}$$

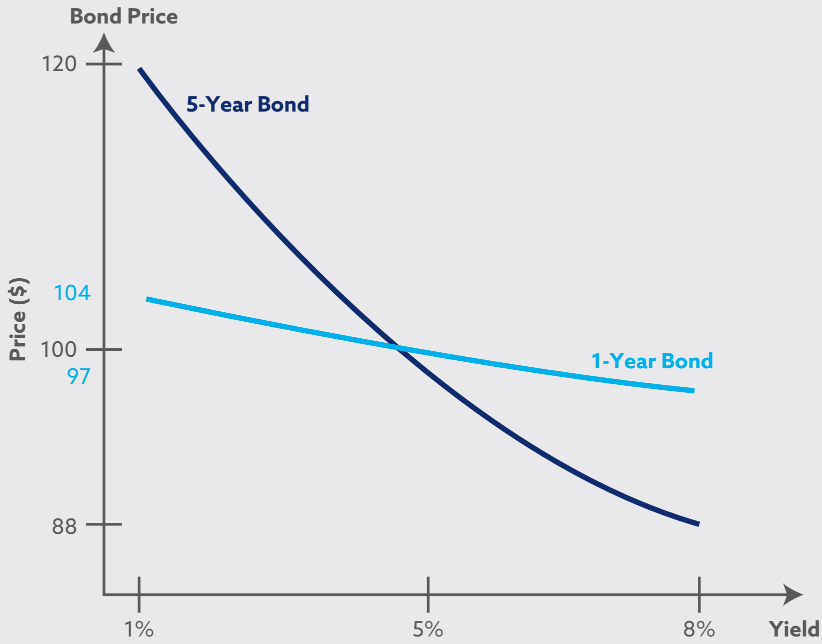
Because longer bonds have a higher modified duration and par yields generally move together (which we will show later), longer-maturity bonds will have more volatile returns than shorter-maturity bonds. Exhibit 7.1 showed that the volatility of returns (the standard deviation of total returns) increases with maturity.

Interest Rate Sensitivity and Modified Duration

Bond prices go down when yields go up, and the prices of longer bonds go down more than those of shorter bonds. **Exhibit 7.2** shows what happens for one-year and five-year bonds when yields change; in this case, both bonds have a 5% coupon. The five-year bond changes more with yields than does the one-year bond, and the slope of the curve measures the degree of sensitivity.

[3] This expression is only an approximation for a number of reasons. First, the percentage change in price is only approximately Modified duration × Change in par yield. Second, the change from one month to the next involves not only the change in yields but also rolling down the curve. A two-year bond this month becomes a shorter one-year and 11-month bond next month. The modified duration we need is some average of the modified duration for a two-year bond and a shorter one-year and 11-month bond, and we must account for "rolling down the curve"—that is, experiencing the return due to the change in yield going from a two-year to one-year and 11-month par yield.

Exhibit 7.2. Price vs. Yield for One- and Five-Year Bonds



Source: Author calculations.

As Exhibit 7.2 shows, longer-maturity bonds are more interest rate sensitive than shorter-maturity ones, but we need a numerical measure for the slope. The interest rate semi-elasticity is the standard measure: the percentage change in price for a unit change in yield, which is negative because the slope is negative (prices go down when yields go up). The elasticity is the slope (in percentage terms) in Exhibit 7.2. The semi-elasticity, also commonly called modified duration (for reasons that will become apparent), is defined as

$$\text{Semi-elasticity (a.k.a. modified duration)} = -\frac{100}{PV(y)} \cdot \frac{dPV(y)}{dy} = \frac{\%\Delta PV}{\Delta y}.$$

Longer-maturity bonds are more sensitive, but by how much? For bonds with a single cash flow, the answer is simple: the bond maturity "modified" by dividing it by $1 + y/2$:[4]

$$\text{Elasticity} = \frac{\text{Maturity}}{1 + y/2}.$$

[4] For single-cash-flow bonds, we use the PV formula, $PV = \frac{CF}{(1+y/2)^{2n}}$, and obtain $\frac{1}{PV(y)} \cdot \frac{dPV(y)}{dy} = \frac{n}{1+y/2}$.
If we measured sensitivity to *continuously compounded* rates (rather than semiannually compounded rates), then the "modification" would disappear and we could simply use the maturity.

For bonds with intermediate coupon cash flows, such as Treasuries, it is still true that longer bonds have greater sensitivity, but the final maturity is, of course, not the right measure of "length" or duration. So, we need to find the appropriate measure. In the early twentieth century, Frederick Macaulay (1938, Chapter 2, pp. 48–49) defined a bond's "duration" as the weighted average amount of time to receiving cash flows, using the present value of payments as weights. He did not realize at the time that this definition turns out to be the correct measure of the bond's length of life to use in calculating the bond's interest rate sensitivity:[5]

$$\text{Semi-elasticity (a.k.a. modified duration)} = \frac{\text{Macaulay duration}}{1 + y/2}.$$

For a single-cash-flow bond, we use the maturity; for other bonds, we use the Macaulay duration.[6] Longer-maturity bonds have a longer Macaulay duration and thus higher sensitivity. Two-, 10-, and 20-year par bonds (in a 5% rate environment and with 5% coupons) have Macaulay durations of 1.9, 8.0, and 12.9 years, respectively.

It is crucial to understand that despite sharing the word "duration," Macaulay duration and modified duration are distinct concepts. Modified duration is an interest rate sensitivity measure, which may apply to a wide range of fixed-income instruments (e.g., options as well as bonds). Macaulay duration measures time (duration in the true sense of the word) and technically applies only to fixed-cash-flow bonds.

---

[5] Bierwag and Fooladi (2006, p. 149) asserted that Macaulay (1938) mentioned, in a footnote, that his duration could be related to the derivative of the bond price, but we have not found that footnote. Macaulay used his duration in discussing the relation between volatility of short-term and long-term *yields* but did not apply his duration to the problem of yields versus prices. To the best of our knowledge, Fisher (1966, p. 113) was the first to provide an explicit derivation of the modified duration relationship for continuously compounded rates. Fisher and Weil (1971) made explicit the use of modified duration for immunization of bond portfolios.

[6] For the technically inclined, the Macaulay duration for a fixed-cash-flow instrument is

$$\text{Macaulay duration} = \sum_{i=1}^{n} t_i \cdot \frac{\text{PV}_i}{\text{PV}_{tot}},$$

where t_i = Time (maturity in years) of cash payment i, PV_i = Present value of cash payment i, and PV_{tot} = Total (sum) of cash flow PVs.

Subsets of the Full Period

Exhibit 7.3 and **Exhibit 7.4** break up the full-period data in Exhibit 7.1 by showing summary statistics for the first half (through the period of rising inflation to 1980) and second half (moderating inflation from 1981 to 2025) of the past 100 years. Income returns were higher during the latter period (because of the higher par bond yields), which pushed up total returns.[7] Furthermore, at the long end, returns in excess of yield (effectively capital gains) were also higher.

Exhibit 7.3. Par Bond Yields and Returns, 1925–1980

	1 M	3 M	6 M	1 Yr	2 Yr	3 Yr	5 Yr	7 Yr	10 Yr	20 Yr	Long
Par bond yield (average)	2.81	2.91	3.05	3.22	3.44	3.60	3.75	3.85	3.91	3.98	4.01
Compound Annual Return											
Total return	2.80	2.94	3.17	3.37	3.58	3.50	3.42	3.26	2.90	2.69	2.59
Income return	2.80	2.88	3.03	3.21	3.44	3.60	3.76	3.86	3.92	3.99	4.02
Return in excess of yield	0.00	0.05	0.13	0.15	0.13	−0.09	−0.32	−0.57	−0.98	−1.25	−1.38
Arithmetic Mean Annual Return											
Total return	2.83	2.98	3.21	3.40	3.61	3.55	3.49	3.34	3.01	2.86	2.80
Annualized Standard Deviation											
Total return	0.79	0.93	1.07	1.47	2.32	3.05	3.90	4.37	5.13	6.26	6.51
Income return	0.79	0.85	0.85	0.82	0.76	0.73	0.70	0.68	0.66	0.63	0.63
Return in excess of yield	0.00	0.30	0.65	1.26	2.20	2.96	3.84	4.33	5.10	6.24	6.49

Source: Author calculations from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in Chapter 8.

[7] Average par bond yields and income returns are not exactly the same, because of compounding: Par bond yields are semi-annually compounded yields, while the annualized income returns are uncompounded—the simple monthly income return times 12.

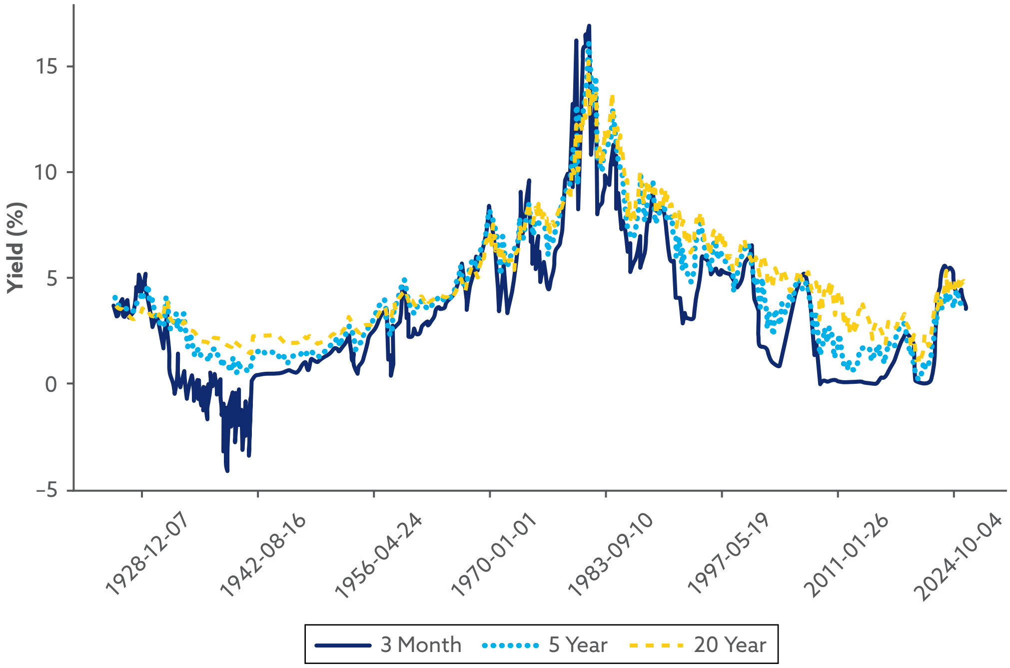
Exhibit 7.4. Par Bond Yields and Returns, 1981–2025

	1 M	3 M	6 M	1 Yr	2 Yr	3 Yr	5 Yr	7 Yr	10 Yr	20 Yr	Long
Par bond yield (average)	3.79	4.03	4.14	4.33	4.56	4.75	5.07	5.32	5.53	5.93	5.91
Compound Annual Return											
Total return	3.82	4.22	4.49	4.84	5.34	5.87	6.60	6.92	7.56	7.62	7.53
Income return	3.82	4.06	4.17	4.36	4.60	4.79	5.12	5.38	5.60	6.01	5.99
Return in excess of yield	0.00	0.16	0.30	0.46	0.70	1.02	1.40	1.46	1.86	1.52	1.45
Arithmetic Mean Annual Return											
Total return	3.87	4.29	4.56	4.92	5.46	6.02	6.84	7.26	8.04	8.50	8.82
Annualized Standard Deviation											
Total return	0.94	1.06	1.20	1.57	2.49	3.44	5.15	6.61	8.13	11.65	13.55
Income return	0.94	0.99	0.99	1.00	1.01	0.99	0.97	0.94	0.91	0.85	0.84
Return in excess of yield	0.00	0.21	0.53	1.09	2.17	3.19	4.96	6.44	7.97	11.49	13.40

Source: Author calculations from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in Chapter 8.

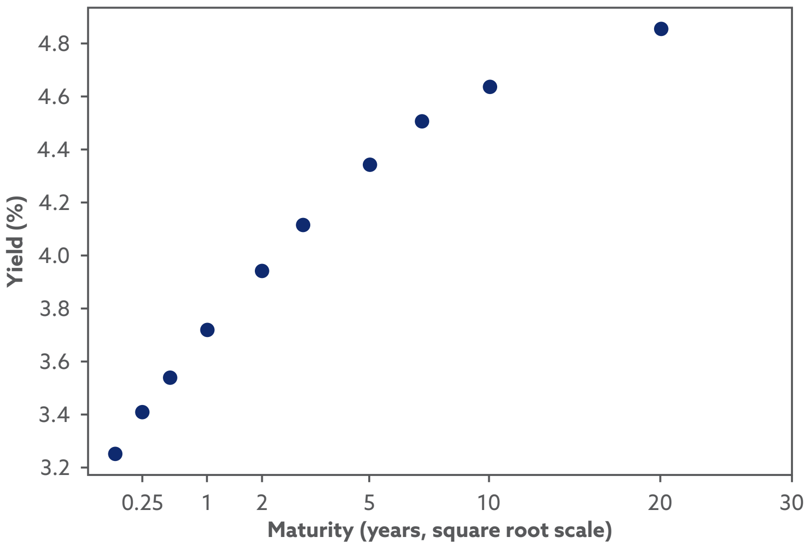
Exhibit 7.5 shows the time series for 2-, 10-, and 20-year yields—a "yield pyramid" of rising yields peaking in 1981 and falling thereafter. **Exhibit 7.6** shows the average par bond curve over the whole period, a graphical representation of the first row of Exhibit 7.1. We use as the horizontal axis the square root of the maturity, simply to expand somewhat the scale at the short end and compress the scale at the long end.

Exhibit 7.5. Par Bond Yields, 1926–2025



Source: Author calculations. CFI par bond yields calculated from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in Chapter 8.

Exhibit 7.6. Average Par Bond Curve, 1926–2025

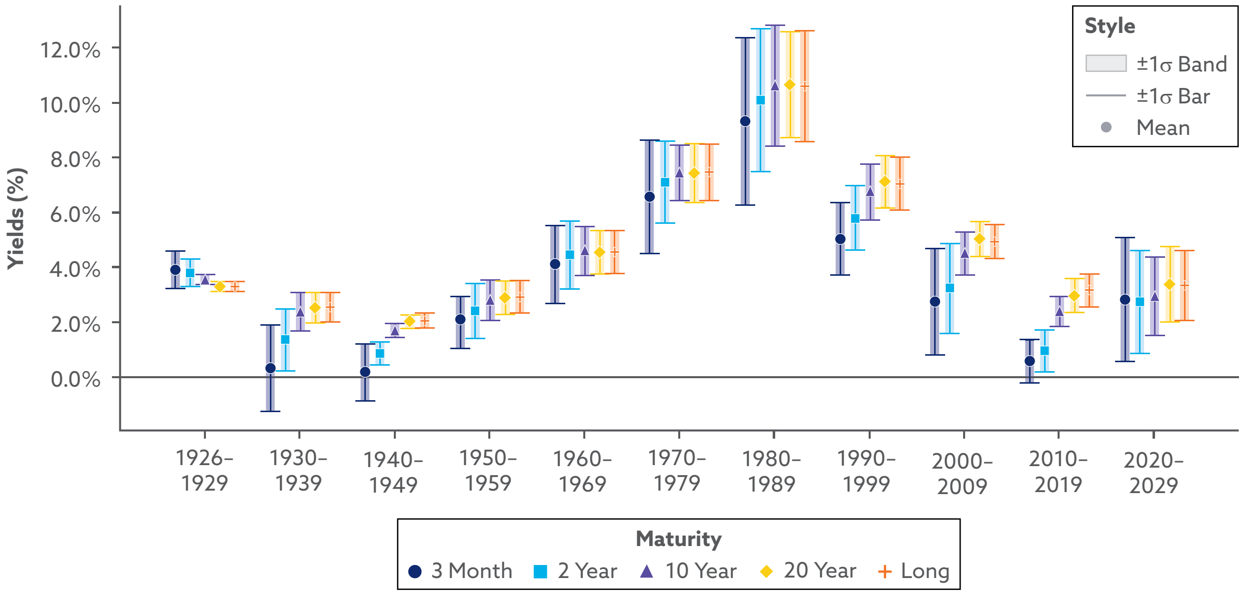


Source: Author calculations. CFI par bond yields calculated from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in the Chapter 8.

Decade-by-Decade Par Bond Yields and Returns

The par bond yield is the yield on a par bond—a bond trading at $100 with coupon equal to yield. **Exhibit 7.7** shows the decade averages for three-month and 2-, 10-, 20-year and longest-maturity par bond yields, together with bars showing the spread around the average. **Exhibit 7.8** is a montage of decade-by-decade average par bond yield curves. Except for the 1926–1929 period and the 2020–25 period (at the short end), the average curves are consistently upward sloping.

Exhibit 7.7. Par Bond Yield Curves by Decade Showing One-σ Bands



Source: Author calculations. CFI par bond yields calculated from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in Chapter 8.

Exhibit 7.8. Average Par Bond Yield Curves, 1926–2025

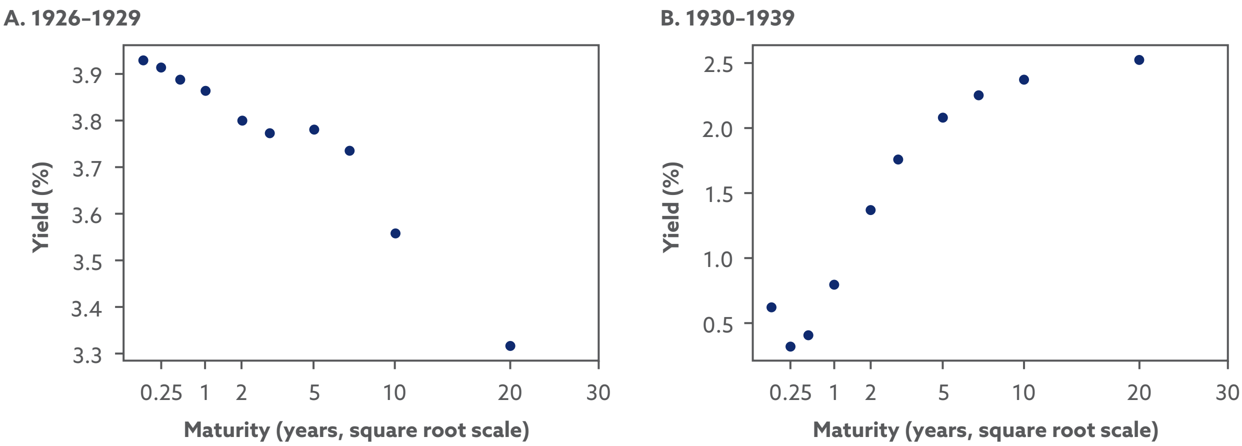
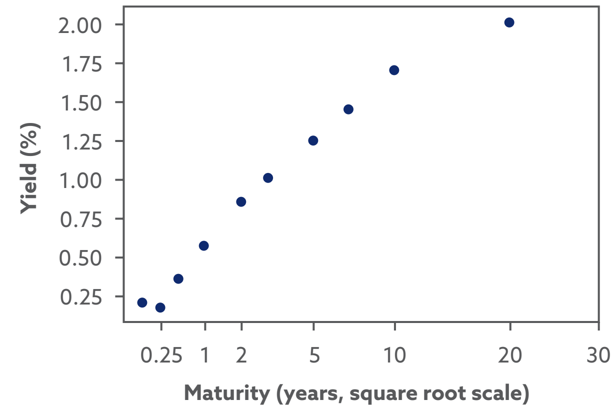
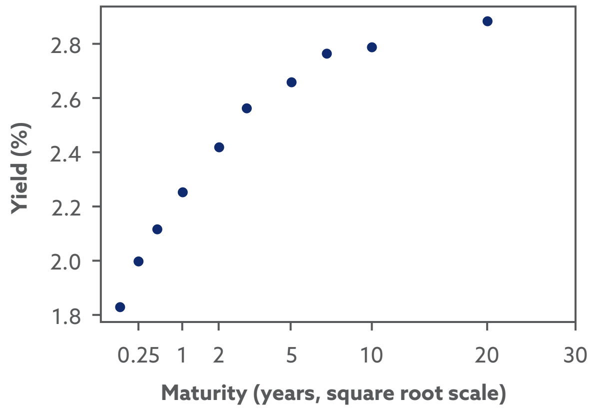


Exhibit 7.8. Average Par Bond Yield Curves, 1926–2025 (*continued*)

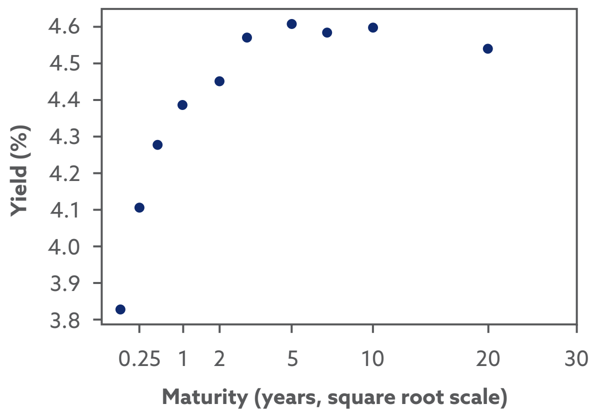
C. 1940–1949



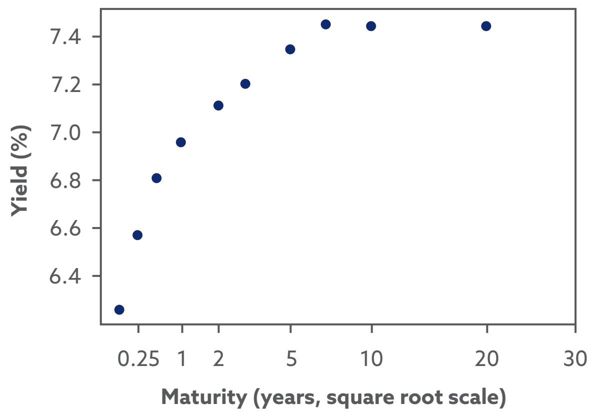
D. 1950–1959



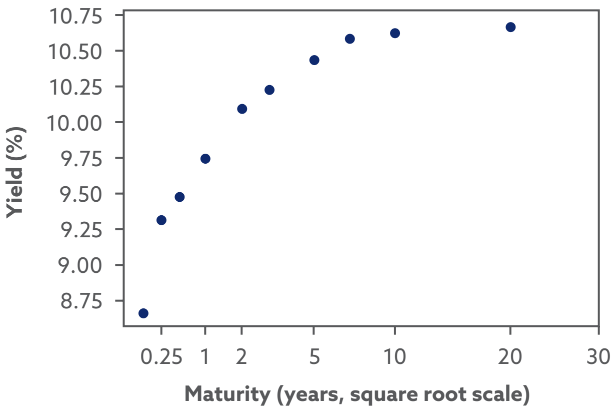
E. 1960–1969



F. 1970–1979



G. 1980–1989



H. 1990–1999

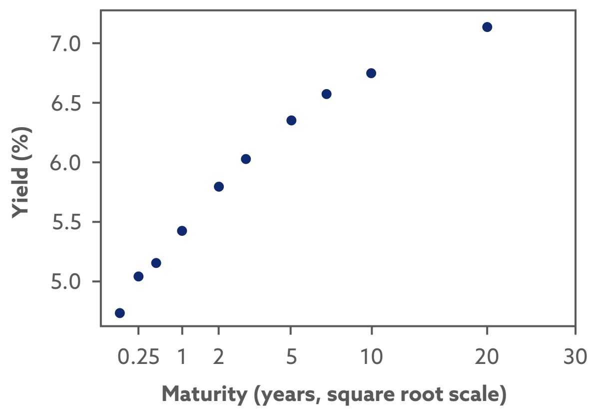
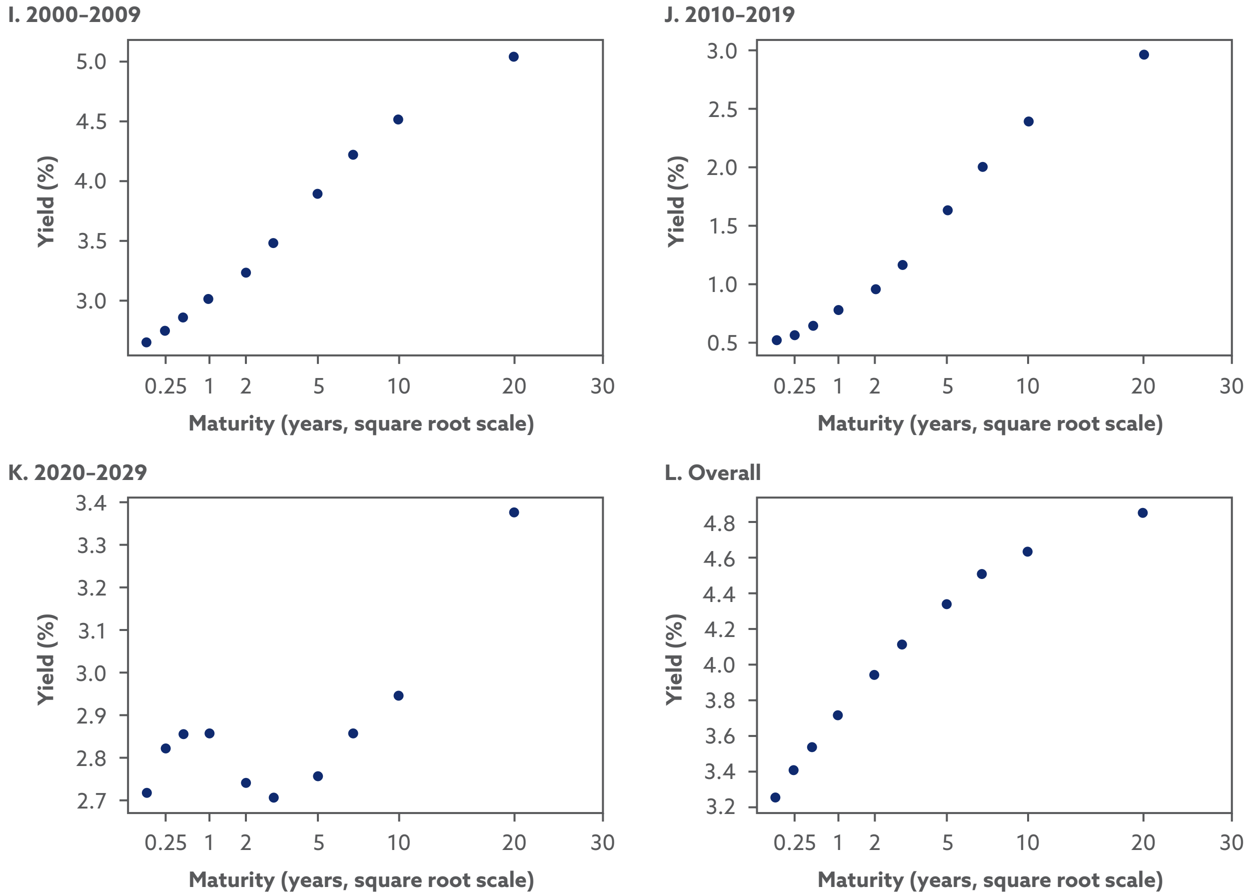


Exhibit 7.8. Average Par Bond Yield Curves, 1926–2025 (*continued*)



Source: Author calculations. CFI par bond yields calculated from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in Chapter 8.

Year-By-Year Par Bond Yields and Returns

Our CFI datasets consist of estimated forward rates covering the full US Treasury bond yield curves for each month-end over the full 100-year period from 1926–2025. From these estimated forward rates, we can construct yield curves of any type, such as par bonds, zero coupon or discount bonds, and annuities. We have chosen to construct par bonds because most US Treasury bonds are issued at $100 with semi-annual coupons. From our CFI estimated forward rates, we can construct the monthly and annualized yields, returns in excess of yields, and total returns for synthetic bonds of any maturity or duration.

Exhibit 7.9 shows par bond annualized yields (last trading day of the year) and par bond total returns (during the year for buying the designated maturity at the beginning of each month and then selling at the end of the month) annually for a few selected maturities, over the period 1926–2025 (starting from 1925 year-end). The exhibit shows the 1-month, 1-year, 10-year, and 20-year maturities.

Exhibit 7.9. Par Bond Yields (end of year, %) and Returns (during the year, %)

	1 Month		1 Year		10 Year		20 Year		Long	
	Yield	Total Return	Yield	Total Return	Yield	Total Return	Yield	Total Return	Yield	Total Return
1926	3.26	3.52	3.26	4.22	3.55	5.23	3.40	6.82	3.40	7.03
1927	3.33	3.43	3.29	3.19	3.25	5.68	3.02	9.06	3.02	9.25
1928	4.73	4.18	4.54	2.52	3.62	−0.19	3.26	−0.63	3.27	−0.74
1929	3.57	4.75	3.51	5.52	3.42	4.72	3.24	3.60	3.23	3.81
1930	2.84	3.24	2.81	3.88	3.09	5.89	3.14	4.85	3.13	4.35
1931	4.13	2.53	4.07	1.56	4.11	−5.12	3.88	−7.32	3.87	−9.34
1932	0.18	0.87	0.34	6.02	2.82	10.46	2.91	14.32	2.94	16.98
1933	0.47	0.28	1.91	0.49	3.07	0.72	2.96	2.10	2.93	2.99
1934	−0.16	0.24	0.15	3.20	2.31	9.91	2.49	10.66	2.53	11.06
1935	1.02	0.19	0.04	0.49	1.83	7.07	2.02	10.49	2.06	12.07
1936	−0.41	−0.27	0.29	0.19	1.68	3.58	1.91	4.04	1.96	4.20
1937	0.01	0.37	0.13	1.45	1.88	0.39	2.12	−1.18	2.15	−1.71
1938	−0.91	−0.57	−0.93	2.06	1.54	5.68	1.85	7.02	1.88	7.32
1939	−0.28	−0.36	−1.16	1.07	1.45	2.93	1.83	2.39	1.85	2.28
1940	−0.18	−0.08	−1.03	0.49	1.26	3.89	1.55	7.07	1.55	7.27
1941	0.05	−1.67	0.26	−0.45	1.35	0.93	1.59	0.93	1.59	0.92
1942	0.35	0.25	0.86	1.01	1.90	2.80	2.25	2.93	2.34	3.02
1943	0.28	0.33	0.79	1.09	1.96	2.08	2.28	2.12	2.35	2.14
1944	0.20	0.30	0.84	0.92	1.94	2.74	2.26	2.80	2.32	2.83
1945	0.12	0.26	0.85	1.14	1.48	6.90	1.83	10.09	1.89	11.23
1946	0.24	0.22	0.87	1.00	1.64	0.70	1.93	0.47	1.96	0.37
1947	0.90	0.42	1.06	0.94	1.99	−0.95	2.26	−3.15	2.28	−3.56
1948	1.14	0.91	1.22	1.01	1.95	2.78	2.17	3.85	2.18	3.97
1949	1.06	1.08	1.09	1.34	1.70	4.61	1.89	7.06	1.88	7.13
1950	1.36	1.19	1.46	0.94	1.94	−0.15	2.08	−0.98	2.06	−0.92
1951	1.72	1.45	1.87	1.28	2.40	−1.69	2.53	−4.78	2.51	−4.31
1952	1.97	1.61	2.02	1.94	2.45	2.18	2.66	0.52	2.62	1.04
1953	1.33	1.77	1.54	2.77	2.34	3.66	2.70	2.60	2.81	4.61
1954	1.04	0.89	1.10	1.56	2.40	1.71	2.45	7.04	2.46	9.52

(continued)

Exhibit 7.9. Par Bond Yields (end of year, %) and Returns (during the year, %) (*continued*)

	1 Month		1 Year		10 Year		20 Year		Long	
	Yield	Total Return	Yield	Total Return	Yield	Total Return	Yield	Total Return	Yield	Total Return
1955	2.46	1.55	2.68	0.83	2.80	−1.25	2.81	−2.62	2.98	−3.95
1956	3.13	2.40	3.44	2.66	3.53	−3.60	3.46	−6.25	3.41	−6.19
1957	2.78	3.02	2.79	4.12	2.85	9.41	3.11	8.87	3.20	8.05
1958	2.37	1.37	3.04	2.17	3.87	−5.09	3.65	−4.49	3.89	−11.37
1959	4.16	2.83	4.87	2.56	4.85	−3.90	4.43	−6.35	4.41	−7.20
1960	2.02	2.51	2.62	5.99	3.68	13.90	3.74	14.12	3.87	16.14
1961	2.53	2.04	3.15	2.78	4.14	−0.10	4.06	−0.47	4.13	−1.12
1962	3.02	2.66	3.04	3.43	3.80	6.90	3.87	6.71	3.93	8.16
1963	3.59	3.08	3.79	2.69	4.12	1.23	4.16	−0.08	4.16	−0.29
1964	3.61	3.53	3.88	3.92	4.02	3.39	4.08	5.13	4.17	4.00
1965	4.54	3.92	4.90	3.22	4.65	−0.83	4.27	1.64	4.50	−1.58
1966	4.76	4.71	4.96	5.15	4.51	5.64	4.45	2.25	4.44	4.71
1967	4.54	4.12	5.73	4.25	5.57	−3.39	5.48	−8.09	5.48	−11.45
1968	6.39	5.28	6.29	5.20	5.98	2.09	5.89	0.36	5.89	−1.51
1969	7.30	6.59	8.18	5.40	7.62	−5.39	6.97	−5.80	7.14	−6.71
1970	4.67	6.39	4.96	10.43	6.48	16.00	6.44	12.94	6.45	14.36
1971	3.43	4.28	4.35	6.07	5.85	11.41	5.86	13.12	6.03	10.78
1972	5.10	3.84	5.59	4.39	6.38	2.52	5.76	6.03	6.01	4.82
1973	7.67	7.07	7.26	5.37	6.84	3.68	7.26	−4.69	7.23	−1.73
1974	6.73	8.13	7.15	8.05	7.06	5.97	7.84	2.06	7.68	3.02
1975	5.20	5.72	6.16	8.47	7.66	3.39	7.92	7.65	7.83	6.56
1976	4.36	5.05	4.88	7.92	6.81	14.57	7.14	17.30	7.05	16.61
1977	5.92	5.15	6.96	4.55	7.77	0.57	7.91	−0.28	7.85	−0.22
1978	9.04	7.25	10.53	5.26	9.15	−0.87	8.89	−1.25	8.98	−1.62
1979	10.74	10.48	11.70	9.02	10.28	1.77	10.00	−0.93	10.09	0.27
1980	13.27	11.51	13.50	10.18	12.39	−1.62	11.90	−4.30	12.05	−3.22
1981	9.98	15.14	13.55	14.73	13.97	4.54	13.65	0.22	13.65	0.36
1982	8.57	10.71	8.83	18.22	10.72	35.69	10.57	41.85	10.57	41.79
1983	9.04	8.84	10.07	9.21	11.80	4.54	11.79	1.22	11.79	1.22
1984	8.15	9.78	9.23	12.88	11.54	14.19	11.62	14.10	11.62	14.12

(*continued*)

Exhibit 7.9. Par Bond Yields (end of year, %) and Returns (during the year, %) (*continued*)

	1 Month		1 Year		10 Year		20 Year		Long	
	Yield	Total Return	Yield	Total Return	Yield	Total Return	Yield	Total Return	Yield	Total Return
1985	6.31	7.57	7.69	10.88	9.09	29.03	9.54	32.02	9.41	34.95
1986	5.07	5.98	6.04	8.60	7.33	22.06	8.06	23.64	7.68	29.76
1987	3.65	5.17	7.29	6.58	8.89	−1.95	9.08	−1.48	8.98	−5.73
1988	6.60	6.16	9.10	6.63	9.16	7.42	9.09	8.89	9.03	8.51
1989	6.50	8.21	7.91	10.11	8.00	17.37	8.09	19.14	8.02	20.83
1990	5.91	7.61	6.96	9.32	8.12	8.15	8.39	5.66	8.26	5.92
1991	3.93	5.42	4.22	9.48	6.84	18.70	7.45	19.22	7.47	18.24
1992	2.91	3.41	3.64	5.34	6.80	8.65	7.44	8.22	7.45	7.99
1993	2.94	2.88	3.65	4.03	5.89	14.72	6.56	17.95	6.50	20.43
1994	4.36	3.81	7.26	2.58	7.89	−6.56	8.01	−7.69	7.89	−9.61
1995	4.68	5.46	5.20	8.31	5.70	25.49	6.06	32.11	6.00	35.33
1996	5.09	5.12	5.58	5.68	6.49	0.93	6.75	−1.08	6.69	−2.90
1997	5.04	5.07	5.57	6.19	5.83	12.21	6.02	16.01	5.96	17.23
1998	4.45	4.73	4.61	6.29	4.82	14.42	5.48	12.79	5.24	16.71
1999	5.10	4.57	6.07	4.41	6.57	−6.68	6.85	−9.54	6.61	−12.34
2000	5.92	5.81	5.51	7.09	5.23	17.49	5.61	22.12	5.49	22.99
2001	1.66	3.75	2.06	7.22	5.17	6.75	5.79	3.41	5.58	4.17
2002	1.15	1.64	1.21	3.20	3.91	17.85	4.94	17.61	4.86	17.19
2003	0.86	1.01	1.24	1.47	4.28	2.49	5.16	2.03	5.11	0.86
2004	1.85	1.20	2.74	0.87	4.27	5.92	4.91	8.71	4.84	9.13
2005	4.05	2.94	4.43	2.30	4.43	3.56	4.63	8.41	4.56	8.71
2006	4.89	4.64	5.01	4.29	4.74	2.74	4.91	0.96	4.81	−0.33
2007	2.84	4.62	3.27	6.07	4.14	10.50	4.53	10.07	4.46	10.62
2008	0.03	1.43	0.40	4.89	2.54	21.20	3.05	27.96	2.66	42.80
2009	0.03	0.10	0.51	0.70	3.90	−6.40	4.60	−17.15	4.61	−27.90
2010	0.07	0.12	0.34	0.73	3.30	10.50	4.19	10.04	4.34	8.84
2011	0.00	0.05	0.16	0.57	1.91	18.73	2.69	30.28	2.92	35.87
2012	0.03	0.05	0.18	0.25	1.76	4.94	2.66	3.46	2.94	2.66
2013	0.01	0.03	0.16	0.29	3.04	−7.40	3.77	−12.66	3.96	−15.09
2014	0.01	0.02	0.28	0.21	2.19	11.80	2.58	23.99	2.78	30.19

(*continued*)

Exhibit 7.9. Par Bond Yields (end of year, %) and Returns (during the year, %) (*continued*)

	1 Month		1 Year		10 Year		20 Year		Long	
	Yield	Total Return	Yield	Total Return	Yield	Total Return	Yield	Total Return	Yield	Total Return
2015	0.10	0.01	0.80	0.22	2.32	1.61	2.81	−0.80	3.03	−2.53
2016	0.41	0.18	0.92	0.83	2.48	0.65	2.90	1.27	3.08	1.65
2017	1.23	0.77	1.80	0.63	2.42	3.31	2.60	8.38	2.75	10.17
2018	2.40	1.79	2.64	1.89	2.70	0.42	2.88	−1.16	3.02	−2.45
2019	1.46	2.12	1.62	3.01	1.93	9.73	2.28	13.03	2.39	16.33
2020	0.05	0.43	0.10	1.87	0.92	11.53	1.43	17.50	1.66	19.20
2021	−0.02	0.04	0.41	−0.03	1.48	−3.04	1.92	−6.17	1.90	−4.29
2022	4.00	1.39	4.71	−0.75	3.85	−16.40	4.19	−27.22	3.99	−32.43
2023	5.56	5.08	4.77	4.88	3.89	3.35	4.22	3.53	4.03	2.40
2024	4.11	5.36	4.20	5.02	4.58	−1.40	4.88	−4.21	4.80	−8.45
2025	3.44	4.03	3.52	4.42	4.18	8.41	4.80	5.88	4.86	3.60

Source: Author calculations. CFI par bond yields and returns calculated from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in Chapter 8.

Key Takeaways

- We extended our earlier CFI work, using spline methods to estimate forward interest rates, which we then used to estimate US Treasury yield curves and synthetic US Treasury bond returns across the yield curves.
- Bond yields generally rose over roughly the first half of the period, reaching a peak above 15% for long bonds in September 1981. Then, yields fell until the early 2020s. These trends created negative returns in excess of yield in the earlier period (1926–1980) and positive returns in excess of yield in the later period (1981–2025).
- Overall, the average income returns were 3.21% for short bonds (one-month maturities) and 4.79% for long bonds (20 years), providing a historical horizon premium of 1.58%.
- Total returns over the 1926–2025 period were approximately the same as the yields. The yield in 1926 was similar to the yield in 2025, so returns in excess of yield are near zero over the whole period.
- Most bond return volatility comes from the change in bond yields and is related to a bond's modified duration or interest rate sensitivity: As yields rise and fall, bond prices fall and rise.

References

Bierwag, Gerald O., and Iraj Fooladi. 2006. "Duration Analysis: An Historical Perspective." *Journal of Applied Finance* 16 (2): 144–60. www.researchgate.net/publication/265146183_Duration_Analysis_An_Historical_Perspective.

Coleman, Thomas S., Lawrence Fisher, and Roger G. Ibbotson. 1989. *U.S. Treasury Yield Curves, 1926–1988*. Moody's Investors Service.

———. 1992. "Estimating the Term Structure of Interest Rates from Data That Include the Prices of Coupon Bonds." *Journal of Fixed Income* 2 (2): 85–116. doi:10.3905/jfi.1992.408048.

———. 1993. *Historical U.S. Treasury Yield Curves*. Ibbotson Associates, Moody's Investors Service.

Fisher, Lawrence. 1966. "An Algorithm for Finding Exact Rates of Return." *Journal of Business* 39 (1): 111–18. doi:10.1086/294845.

Fisher, Lawrence, and Roman L. Weil. 1971. "Coping with the Risk of Interest-Rate Fluctuations: Returns to Bondholders from Naive and Optimal Strategies." *Journal of Business* 44 (4): 408–31. doi:10.1086/295402.

Macaulay, Frederick R. 1938. "The Concept of Long Term Interest Rates." In *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields and Stock Prices in the United States Since 1856*, 24–53. National Bureau of Economic Research.

CHAPTER 8. TREASURY BOND YIELD AND RETURN METHODOLOGY

Thomas S. Coleman, PhD
Senior Lecturer and Executive Director, Center for Economic Policy, Harris School of Public Policy, University of Chicago

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

We seek to estimate the yields and returns for US Treasury bonds, notes, and bills over 1926–2025. We use a method that first extracts monthly forward-rate yield curves from actual US Treasury security prices and then, from these yield curves, creates synthetic securities based on the work of Coleman, Fisher, and Ibbotson (1989, 1993). We began this effort in collaboration with the late Lawrence Fisher in the 1980s. In recognition of this partnership, we call the results of our study, presented in Chapter 7 and discussed in this chapter, the Coleman Fisher Ibbotson (henceforth CFI) database and methodology. Treasury yield curves are useful not only for understanding the Treasury market itself but also for pricing most other US dollar–denominated bonds based on spreads over the Treasury curve.

The resulting bond data include three components:

- Monthly estimated US Treasury forward rates (yield curves)
- Par bond riskless yields for selected maturities (one, three, and six months; 1, 2, 3, 5, 7, 10, and 20 years; and the longest maturity) and monthly returns for those maturities
- The ability to calculate (using the estimated yield curves) the present value (PV) for any arbitrary US dollar "riskless" (nondefaultable) cash flows and instruments. For example, one can calculate yield curves for par bonds, zero-coupon bonds (zeros), or annuities or price any set of riskless streams of US dollar cash flows over the last hundred years.

To produce the estimated forward rates and par bond yields, we use the underlying bond prices in the CRSP bond data file, which was itself created by Fisher at the University of Chicago. The CRSP data contain prices and characteristics for all market-traded bonds from December 1925 to the present. Collectively with Fisher, however, we recognized that translating the raw prices into forward yield curves could enable the data to be used to value all types of standardized bonds (e.g., par bonds for selected maturities) and any arbitrary cash flows (as long as the cash flows are riskless). Here we expand on our earlier work, offering refinements in the splines estimation techniques, taxation adjustments, and probabilistic maturity adjustments based on callability terms.

 •

Goals and Challenges in Estimating Bond Returns

Our goals for US Treasury markets are straightforward: to estimate yield curves and then returns for selected single-bond portfolios with maturities of, say, three months, 2 years, and 10 years.[1] Calculating a return requires "buying" a bond today (say, with a two-year maturity) and "selling" it one month later (when it has one year and 11 months maturity). To be consistent across time, we wish to buy today an exact two-year bond trading at par ($100) and sell the same bond in one month, when the bond will be one month shorter. Using the actual bonds traded is one approach, but generally there will not be exact-maturity bonds for the maturities we want; so, some interpolation method is necessary for calculating synthetic bonds and portfolios.

The approach we take is to estimate a yield curve for each month and, from that estimated curve, price a synthetic two-year bond. Then, a month later, we estimate the new curve and use that new curve to price the original bond (shorter by one month). The benefit of using estimated yield curves for calculating synthetic bonds is that once we have the curves, we have the flexibility to estimate returns across arbitrary bonds (and portfolios) and for different holding periods—we can estimate a three-month return as easily as a one-month return.

As we have emphasized, par bond yields are the most common (and most intuitive) way to display and describe the yield curve. In Chapter 7, we provide exhibits illustrating par yields for a discrete set of points (one month, three months, etc.). The Federal Reserve publishes the H.15 constant-maturity yields for a set of discrete maturities. It is instructive to think of bond returns in terms of these par bond yields and changes in the yields. We can think heuristically of the total return for a two-year bond from January to February being the sum of multiple components:

- Accrued income (earnings from the promised yield over the month)
- Change in the par bond yield for the on-the-run bond: ModDur × $(-\Delta y)$, with Δy being the change from January to February for the on-the-run bond
- Return from "rolling down" the yield curve, from two years to one year and 11 months, ModDur × $(-\Delta y)$, where now the change is the two-year par bond to the one-year and 11-month par bond for a fixed curve (February)

This decomposition is useful because it highlights why par bond yields for discrete maturities (e.g., 1 year, 2 years, 10 years, such as we provide in Chapter 7 or the Fed publishes in its H.15 data releases) are insufficient for calculating monthly returns. The first and second items, (accrued income and change in the par bond yield) can be approximated from the discrete par bond yields (ours or those from H.15). For a good estimate of the third item (the return from rolling down the yield curve), however, we require more detail: the shape of the curve between 23 months' and 24 months' (two years') maturity. The discrete points in the exhibits in Chapter 7 or the Fed's H.15 data provide too little detail on the shape at and just below two years' maturity. The return resulting from the roll-down can be quantitatively important. The yield curve is generally upward sloping (see Exhibit 7.8), so the return from roll-down will generally be positive. For the decade 2010–2019, the average one-month yield roll-down at a 2-year maturity was

[1] Our detailed tables include a wider range of maturities: one, three, and six months; 1, 2, 3, 5, 7, 10, and 20 years; and the longest maturity. The three maturities quoted in the text are sufficient for illustrating all the issues.

1.5 bps and at 10 years was 1.0 bp. These amounts are not large, but they accumulate over a year to contribute roughly 0.4% and 1%, respectively, to the annualized average return.[2]

The need to calculate synthetic prices for both on-the-run exact-maturity bonds (say, 2 years and 10 years) and the shorter bonds (1 year and 11 months and 9 years and 11 months) forces us to calculate curves that both are flexible (allow variable shapes) and use the market data that are available. The returns we report incorporate that detailed information on the shape of the curve, detail beyond what is available in our par bond tables (or the Fed's H.15 data), which show only discrete points.

The piecewise twisted forward curve methodology we use, discussed in the next section, allows for a flexible curve that can use the rich data available in the markets. **Exhibit 8.1** shows the instantaneous forward rates for December 2021 in Panel A and the resulting continuous par bond curve (blue line) plus the actual bonds (dots) in Panel B. We might consider this line to reflect a typical curve—upward sloping at the short end but flattening somewhat at the longer end. Allowing jumps in the instantaneous forward rates has the benefit that the forward curve is localized: The short and long ends are connected through the relative prices of short- and long-dated bonds, not through the mathematical constraints that are often imposed so that the estimated forward curve is continuous.

Fitting the Treasury curve across time requires a curve methodology that is not only fine grained but also flexible and capable of fitting the wide variety of yield curves actually observed. **Exhibit 8.2** shows par bond curves for December 2022 and 2023. The level and shape of the curves are quite different from each other and from the 2021 curve. In 2022, the curve is steep at the very short end (out to roughly 1 year), inverted from roughly 1.5 to 10 years, and then humped. For December 2023, the curve is fully inverted between 1 and 5 years, is flat between 5 and 10 years, and has a hump at the long end.

Forward Curve Methodology

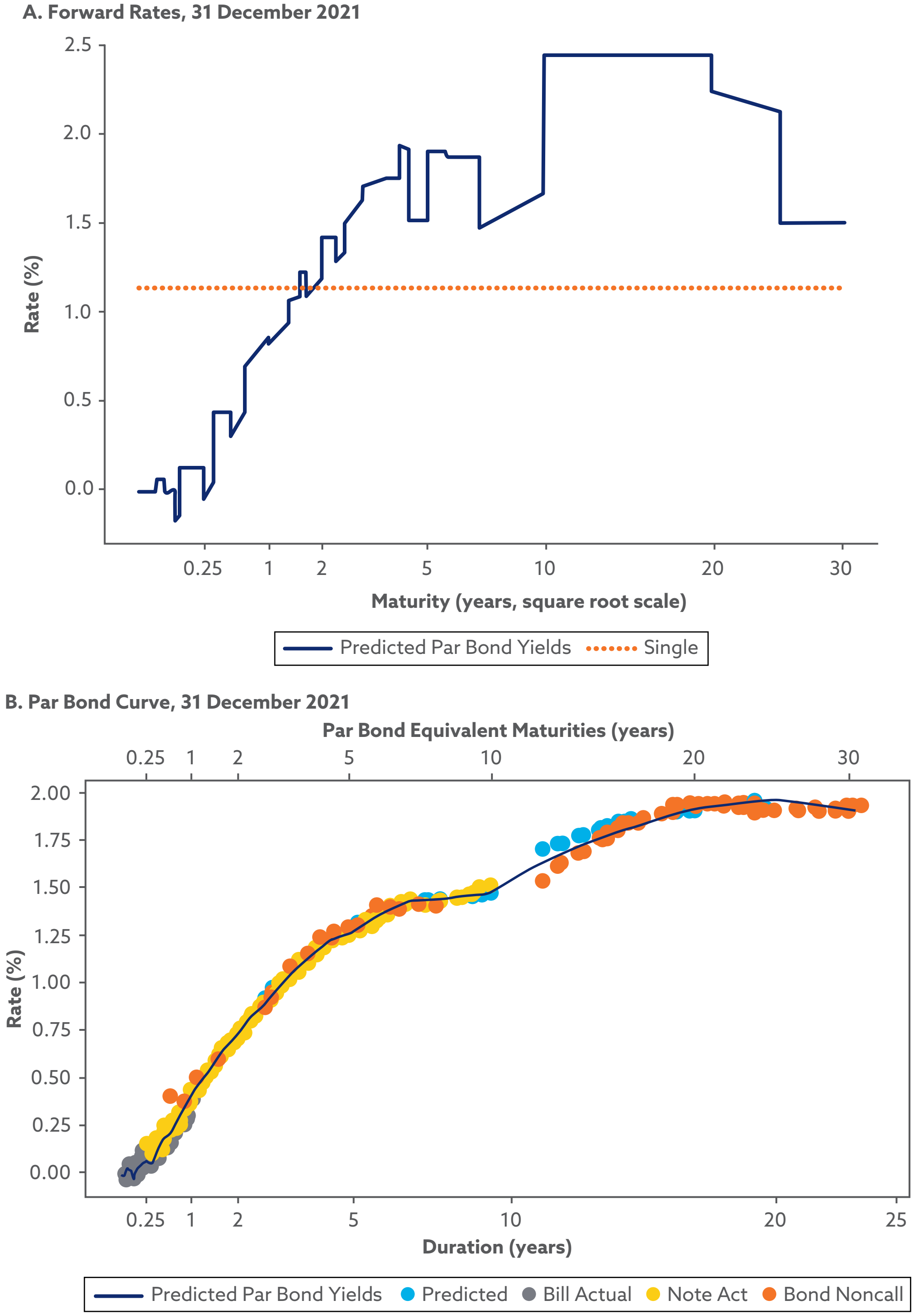
The general approach we use to fit forward curves is split into three components:

- Choosing a parametric functional form for the forward curve: The parametric forward curve translates into a function for the discount factor function for each maturity, $d(t)$. With this discount factor function, we can then calculate the PV of any set of cash flows: $P_0 = \sum_t d(t)C_t + d(T)M$.
- Choosing market data—the bonds—and appropriately describing their cash flow dates and amounts
- Defining and implementing an appropriate objective function and fitting methodology that specifies how to use the market data to fit numerical values for the parameters of the forward curve

Exhibit 8.1 shows an example of the fitted forward curve. For the parametric functional form, we start from the simplest one: instantaneous forward rates that are piecewise constant—that is, assuming that the forward curve is constant between specified breakpoints. Jumps in the

[2] The differences in yield of 1.5 bps and 1.0 bps for the 2- and 10-year bonds, respectively, translate to larger returns because the yield difference needs first to be annualized (multiplied by 12) and then multiplied by the negative duration of the bond.

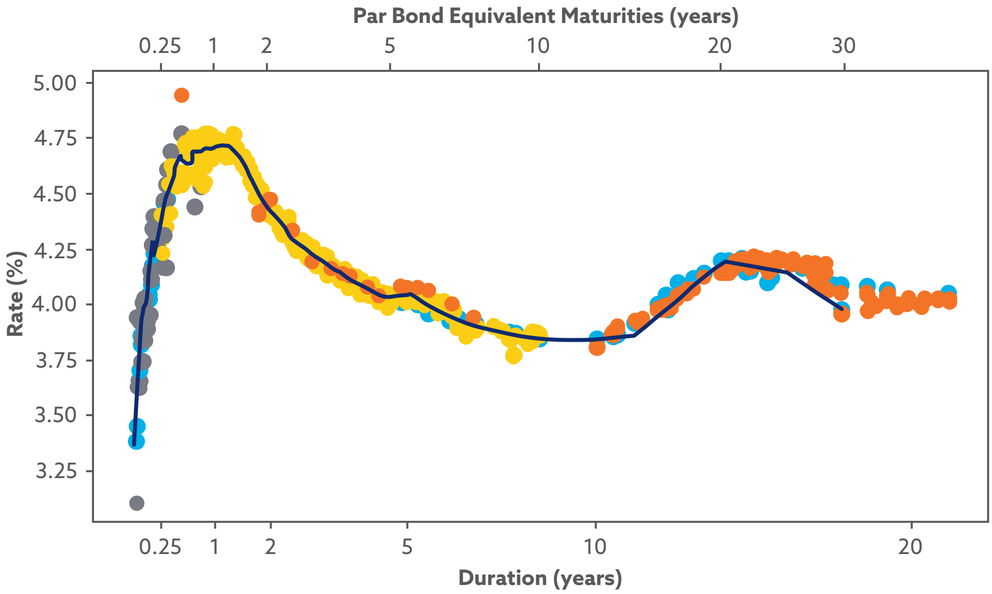
Exhibit 8.1. Forward Rates and Par Yield Curve for 2021



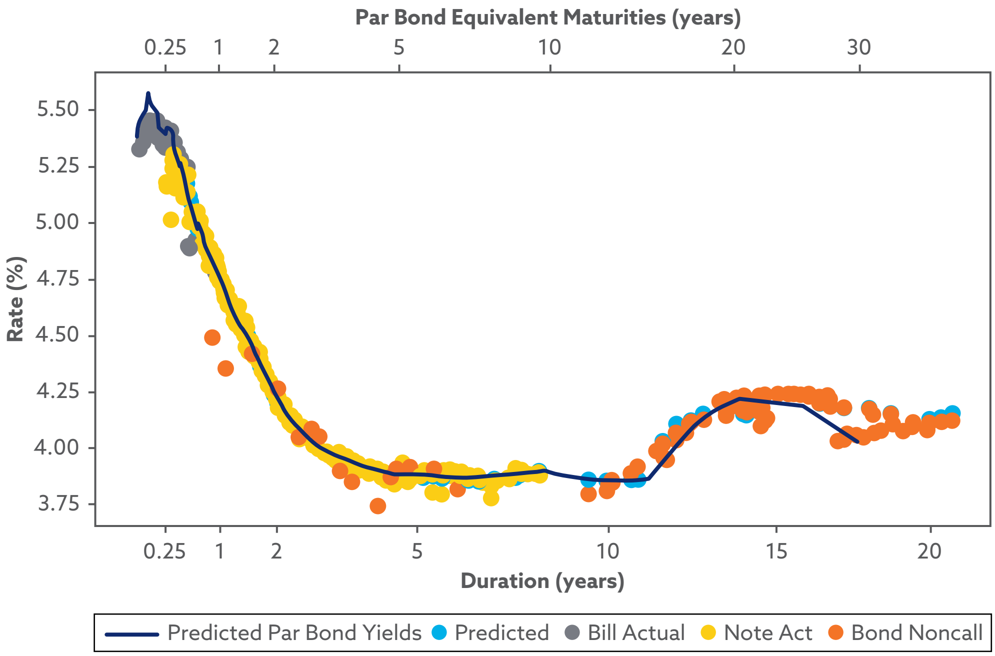
Source: Author calculations. CFI par bond rates and actual versus predicted yields calculated from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in the text.

Exhibit 8.2. Par Yield Curves for 2022 and 2023

A. Par Bond Curve, 30 December 2022



B. Par Bond Curve, 29 December 2023



Source: Author calculations. CFI par bond rates and actual versus predicted yields calculated from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in the text.

instantaneous forward rates do not cause problems with the zero rates or discount factors. As we show in the sidebar, the discount factor and zero-coupon yields (hereafter, zero yields) are functions of the *integral* of the forward curve and thus are always continuous, with jumps in forward rates corresponding to changes in the slope of discount factors or zero rates.[3]

We start with the following breakpoints:

- *Short end:* 7 days, 14 days, 21 days, 28 days, 0.32 year, 0.5 year, 1 year
- *Long end:* 2 years, 3 years, 5 years, 10 years, 15 years, 20 years, 25 years, 32 years

For recent decades, these breakpoints roughly correspond to weekly bills at the short end and on-the-run points for notes and bonds. We then split those periods—entering a new breakpoint—when there are many bonds (at least 25 within a break period and 5 in each of the new periods). As Exhibit 8.1 and Exhibit 8.2 show, this method produces a fine-grained curve when there are many bonds, as occurred in the 2020s. We also allow the forward rates to twist, which further smooths the forward curve without inducing undue oscillation.

Yields and Forward Rates

To properly explain forward rates, we need to get technical. A forward rate is the discount rate or interest rate between two dates in the future. We write the price or PV of a bond today as the future cash flows discounted back to today by a *yield-to-maturity*, discussed earlier:

$$PV = \frac{c/2}{1+y/2} + \frac{c/2}{(1+y/2)^2} + \cdots + \frac{c/2}{(1+y/2)^{2n}} + \frac{100}{(1+y/2)^{2n}}.$$

In addition, when we are considering many bonds trading in the same market (as with the US Treasury market), we want to allow a different yield for different maturities. In its simplest form, this is called the zero yield, z_i, which is different for each maturity. Using the zero yield, we now have

$$PV = \frac{c/2}{1+z_{0.5}/2} + \frac{c/2}{(1+z_1/2)^2} + \cdots + \frac{c/2}{(1+z_n/2)^{2n}} + \frac{100}{(1+z_n/2)^{2n}}.$$

Instead of a separate zero rate for each period (a rate that goes from period *i* back to today), we could also write this in steps, going from today to 0.5, from 0.5 to 1, and so on:

$$PV = \frac{c/2}{1+f_{0,0.5}/2} + \frac{c/2}{(1+f_{0,0.5}/2)(1+f_{0.5,1}/2)} + \cdots + \frac{100}{(1+f_{0,0.5}/2)\cdots(1+f_{n-0.5,n}/2)}.$$

[3] It would seem very natural to generalize piecewise constant to piecewise linear forward rates, continuous across breakpoints but with different slopes. Doing so, however, would produce very poor results. As discussed in Hagan and West (2015), the piecewise linear approach produces oscillations in forward rates and is very nonlocal.

Here, $f_{0.5,1}$ is the *forward* rate or forward yield from ½ year to 1 year, and $f_{n-0.5,\,n}$ is the forward rate from n – 0.5 years to n years. This is a particularly convenient form for fitting and estimating a yield curve because it allows better control and analysis of the yield or forward curve. In our original work (Coleman et al. 1989, 1993), we assumed forward rates were constant between fixed maturity points. Here, we allow the forward rates to twist, and we more flexibly determine the breakpoints, depending on the number of bonds available.

Actual bonds have payments at different dates, not just on half years, and we need to work with a forward rate function $f(u)$ defined over all future maturities u. We need to define the relationships between the forward curve, the discount curve, and the zero interest rate curve:

$$d(t) = \exp\left[-\int_0^t f(u)du\right],$$

$$z(t) = \frac{\left[\int_0^t f(u)du\right]}{t}, \text{ and}$$

$$d(t) = \exp[-z(t) \cdot t],$$

where

$f(u)$ = forward rate function (continuously compounded)

$d(t)$ = the discount factor function for cash at time t, the PV of \$1 due at time t

$z(t)$ = continuously compounded zero rate from today to t

t = maturity from today, so today = 0

If t is measured in years, $f(u)$ and $z(t)$ are both continuously compounded annual rates.

Now the PV (price plus accrued interest) of any bond can be expressed as

$$P_0 = \sum\nolimits_t d(t)C_t + d(t)M,$$

where

P_0 = price plus accrued interest at time zero

C_t = coupon payment due at time t

M = face amount of bond (usually \$100) due at time T

The price of any bond is now a function of the forward curve, $f(u)$.

Displaying Monthly Par Bond Yields—Macaulay Duration

In Chapter 7 and in the detailed exhibits available in the online appendix, we show annual and month-by-month yields. In the tables, we display the par bond yields for a range of maturities from one month to 20 years. In the graphs, we display both par bond yields *and* the actual-versus-predicted yields for the bonds used in estimation, as in Exhibits 8.1 and 8.2. These graphs show our estimated par bond curve (the blue line), the predicted yields for each bond (light blue dots), and the actual yields for each bond (dots of different colors to denote bills versus notes and bonds).

Each graph contains a wealth of information: the shape of the par bond curve, the quality of fit (any separation between the light blue dots and dots of other colors), and the variety of bonds in the market each month. These graphs also illustrate some of the challenges we face: sharply different levels of yields (for example, roughly 10% in 1978 compared with 4% in 2025); sharply different yield curve shapes (for example, sharply rising at the short end, generally upward sloping in 2021, and inverted in 2023); and challenges with callable bonds containing embedded options (for example, in 1978).

The presence of callable bonds for most of the first 80 years of our sample period compels us to make a slightly unusual choice for our graphs: displaying the yield versus Macaulay duration (using the Macaulay duration for the horizontal scale). For callable bonds, there is really no other good choice, because a callable bond's "maturity" will be some probabilistic combination of time to first call and final maturity. We can, however, use the concept of Macaulay duration (a present-value-weighted measure of the "length" or average time to receipt of cash flows of a bond, in years) and the relation between Macaulay duration and elasticity or modified duration (interest rate sensitivity) to define an option-adjusted Macaulay duration. We first calculate the interest rate sensitivity (modified duration) of the callable bond, which will be lower when the bond has a high probability of call (behaves like a bond to first call) and higher with a low probability of call (a bond to final maturity). We then use the relation

$$\text{ModDur} = \frac{\text{MacaulayDur}}{1+y/2}. \tag{8.1}$$

This "option-adjusted" Macaulay duration, measured in years, allows us to graph both callable and noncallable bonds on a consistent and comparable basis. As an extreme example, the 4.25s of 15 June 1947 (callable 1932) were trading right at par in January and February 1929. First call was 3.3 years, final maturity 18.3 years, and the Macaulay duration roughly 6 years. The Macaulay duration of 6 years for this callable bond corresponded to a par bond with maturity of roughly 6.7 years, even though the final maturity was almost three times that.

Estimation Challenges with Historical Curves—Adapting Forward Curve Methodology

In estimating a yield curve and returns consistently across history, there are a variety of challenges. For the current Treasury market, most of these challenges do not arise: There are many bonds (on the order of 450 bonds outstanding in 2024), they are well distributed across maturities, and all are standardized with no quirks, such as callability. For historical curves, however, this is not the case. To estimate consistent returns, we face these challenges:

- *Curve methodology:* A flexible but robust curve methodology that can be applied across the past 100 years
- *Callability:* The US Treasury issued long-dated callable bonds for most of the period 1925–1984. Through 1955, virtually all long-dated bonds (maturity of 15+ years) were callable, but the portion fell into the 25% range by the 1960s and then rose back to 90% by 1980. **Exhibit 8.3** shows that in 1993, callable bonds still represented more than 35% of bonds with maturities of 15+ years. Any estimation of the long end of the curve must deal with callability.
- *Flower bonds:* Through the 1970s, virtually all long-dated bonds (maturity of 15+ years) were "flower bonds" that could be redeemed at par for payment of estate taxes. This meant that during a period of high interest rates, when low-coupon bonds would normally trade at a deep discount, such a bond would have value above and beyond the discounted value of future cash flows because it could be put (sold) back to the government at par if the owner died.
- *Tax status:* During the 1920s, the Treasury issued mainly partially tax-exempt securities. During the 1930s, it issued mainly fully tax-exempt securities. Starting in the 1940s, it issued fully taxable securities.
- *Sparse data:* In the early years of our sample, there were only a handful of bonds and not very many at the long end. For example, in January 1926, there were 15 usable partially tax-exempt bonds, with one maturing in 1952 and another in 1954—1926 and 28 years' maturity, respectively. But these had calls in 1947 and 1944, which effectively shortened their duration.

Sparse Data, Rich Data, and Choice of Breakpoints

We start with breakpoints, shown previously, that for recent decades roughly correspond to weekly bills at the short end and on-the-run points for notes and bonds. For many of the early decades of our sample, the data are quite sparse. We collapse break periods when there is only one bond in a period. For example, for January 1929, the breakpoints and number of bonds are as follows:

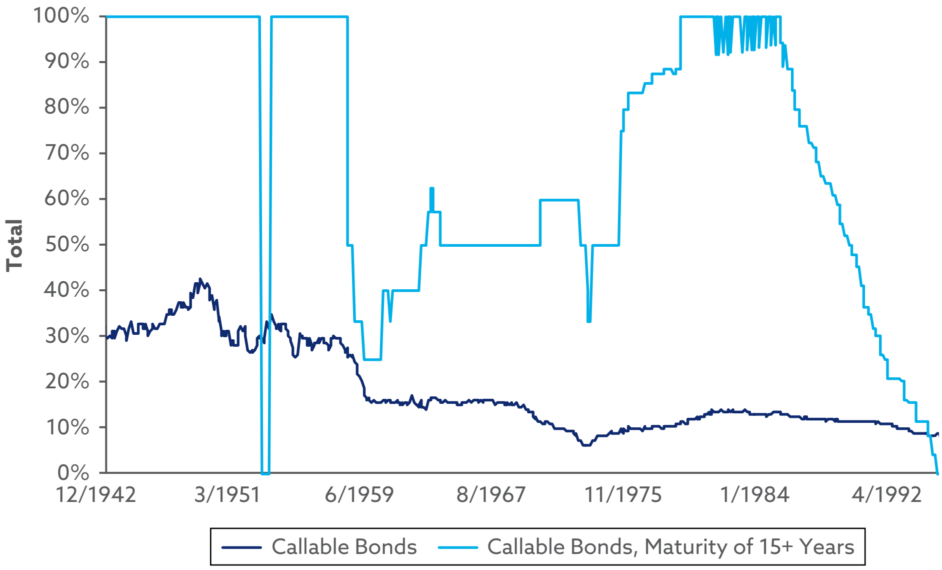
Breakpoints	1–321 days	3.67 years	6.61 years	20.40 years
Number of bonds	4	3	2	5

Callable bonds make the issue considerably more difficult because they are some probabilistic combination of short dated (to first call) and long dated (to final maturity). When there are callable bonds, we use the option-adjusted Macaulay duration to set breakpoints. We first calculate the option-adjusted modified duration using a single rate (flat rate for the whole curve). We then convert the modified duration to an option-adjusted Macaulay duration and then to an equivalent maturity for par bonds (using the estimated single-rate yield). We assign bonds to breakpoint periods (and count the number of bonds in a period) based on their maturity—the actual maturity for noncallable bonds and this "par bond equivalent maturity" for callable bonds.

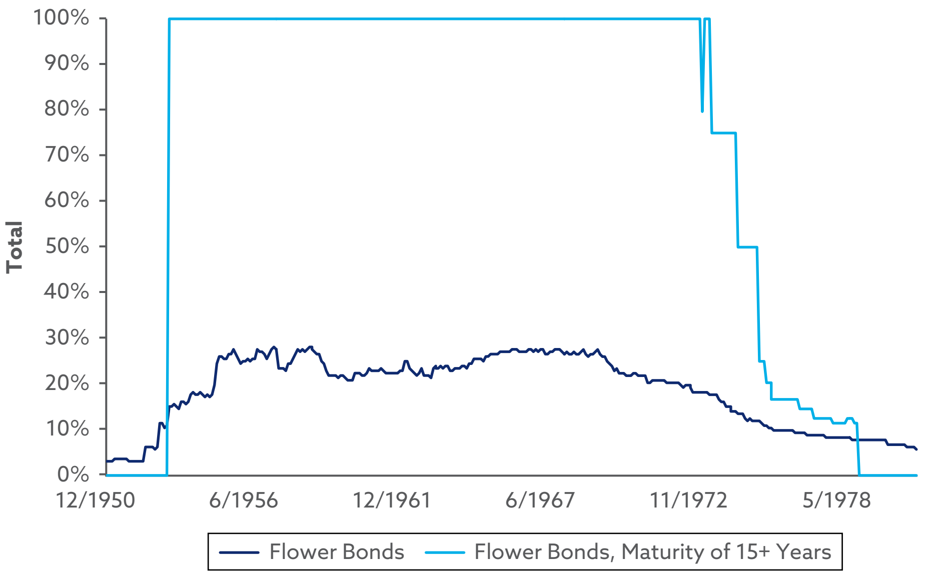
For recent decades, we have many bonds (more than 350 for the 2020s), allowing us to expand breakpoints. When a break period has more than 25 bonds, we split the period but ensure that both the earlier and later periods contain at least 5 bonds.

Exhibit 8.3. Percentage of Callable and Flower Bonds

A. Callable Bonds, December 1942–December 1994



B. Flower Bonds, December 1950–December 1981



Source: Author calculations based on CRSP Treasury bond data.

Tax Status

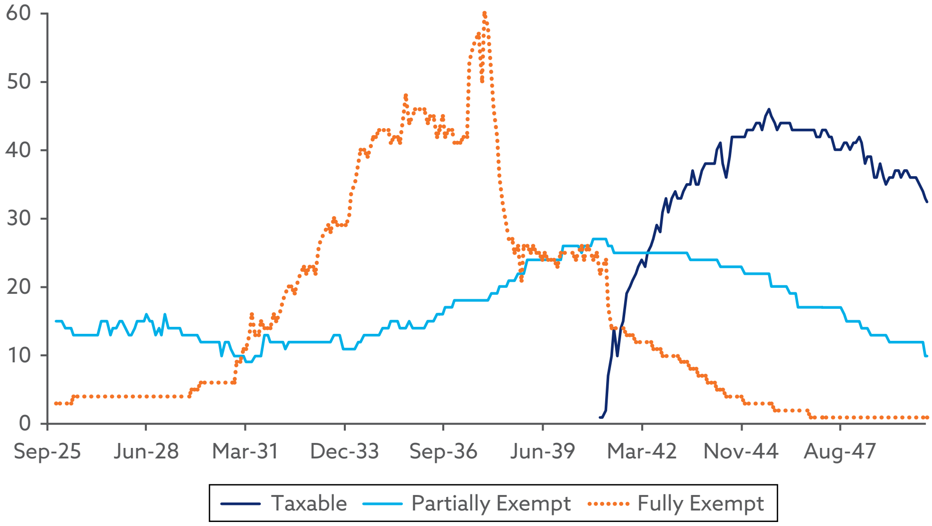
During the 1920s and 1930s, both wholly and partially tax-exempt Treasury bonds were outstanding. Fully taxable bonds appeared in 1940, and all three varieties were being traded until the early 1960s. Among bonds that were similar in other respects, the wholly tax-exempt bonds had the lowest yields and the fully taxable bonds had the highest.

Exhibit 8.4 shows the three regimes. Panel A shows the number of bonds and bills of each type. Panel B analyzes the short end, showing the percentage of all bills (by count) that are in each tax status.

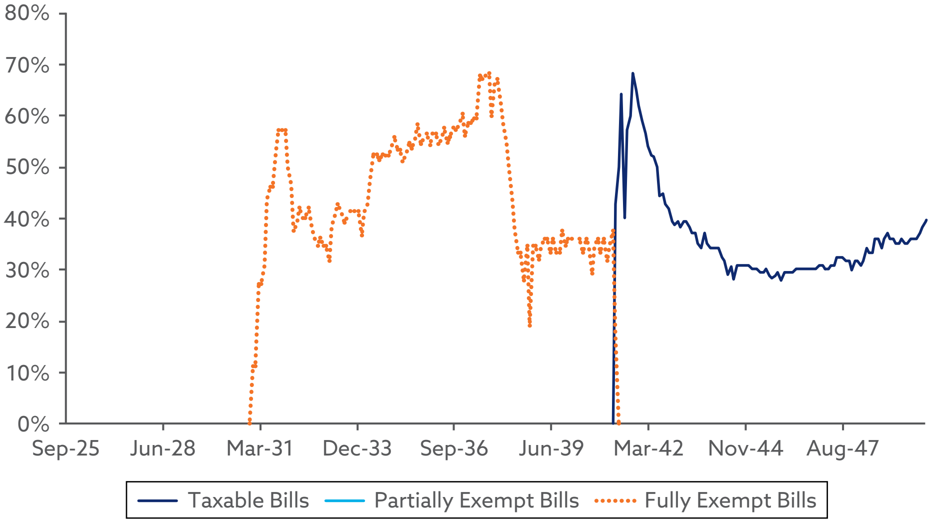
Because yield does depend on tax status, our term structure estimates treat each class separately. Our estimates are not very reliable, however, unless the class includes both short-term and long-term bonds.

Exhibit 8.4. Bonds and Bills by Tax Status, December 1925–December 1949

A. Number of Bonds and Bills by Tax Status



B. Percentage of Bills by Tax Status



Source: Author calculations based on CRSP Treasury bond data.

Exhibit 8.5. Curve Regimes

	Estimated Curves	Curve Overlap	Bond Returns	Return Overlap
Partially exempt	Dec. 1925–Dec. 1932		Jan. 1926–Dec. 1931	
Fully exempt	Jan. 1930–Dec. 1942	Jan. 1930–Dec. 1932	Jan. 1932–Dec. 1941	Dec. 1931–Jan. 1932
Fully taxable, flower	Jun. 1941–Dec. 1973	Jun. 1941–Dec. 1942	Jan. 1942–June 1973	Dec. 1941–Jan. 1942
Fully taxable	Jan. 1973–Dec. 2025	Jan. 1973–Dec. 1973	Jul. 1973–Dec. 2025	Jun. 1973–July 1973

We therefore estimate three separate sets of yield curves:

- **Partially tax exempt, 1925 through 1932:** Before 1930, we have no real choice except to fit a partially exempt curve, although this does not include any short bills.
- **Fully tax exempt, 1930 through 1942:** Starting in 1930, we can estimate a fully exempt curve. In 1931, the long ends of the curves are not too different but, of course, the short ends are different because there are no partially exempt bills. For December 1935, again, the long ends are not much different in shape but the short ends are quite different—in fact, the rates for the short-end fully taxable curve are negative, with bills slightly positive and short-dated notes negative.
- **Fully taxable 1940 through 2025:** Beginning in April 1941, there was a variety of both short- and long-dated taxable bonds, enough to fit a curve. Initially, there are no short-end bills, but by June 1941, there is at least one bill.

To produce consistent returns, we overlap the regimes by a month and calculate the returns as shown in **Exhibit 8.5**.

Callable Bonds and European Option Methodology

The US Treasury issued long-dated callable bonds for most of the 1925–1984 period. Panel A of Exhibit 8.3 shows the fraction of all bonds from 1942 to 1994 that were callable, and we can see that for long periods, callables made up all or the majority of bonds with 15+ years to maturity. The PV of such bonds cannot be represented as a simple sum of discounted cash flows. We approximate callable bonds by evaluating the option exercisable on the first (or next) call date. The PV of such a callable bond will then be

$$PV_c(\text{FwdCrv}) = PV_{nc}(\text{FwdCrv}) - \text{Call}(\text{FwdCrv}), \tag{8.2}$$

where

$PV_{nc}(\text{FwdCrv})$ = PV of noncallable bond to final maturity date as a function of the forward curve

Call(FwdCrv) = European call option to first (or next) call date, exercisable into a bond from the first call to final maturity date, calculated as a function of the forward curve

PV_c(FwdCrv) = resulting PV of callable bond

The noncallable bond and the option are valued off the same forward curve. We value the European option using option methodology based on Black and Scholes (1973), modified for valuing bond options (as discussed in the online methodology appendix). We generally use a normal Black–Scholes-type yield option valuation method (assuming yields are normally distributed) because it allows rates to go below zero, which may be important for periods of ultra-low rates, such as during the Great Depression and after 2008 (when a small number of short-dated callable bonds was left over from issuance during the 1980s). Although this is an approximation because callable Treasuries are generally callable at each coupon date—they are Bermudian rather than European options—with this methodology, we can estimate a noncallable forward curve.

In the online appendix, we discuss the valuation of callable bonds in more detail. Here, we focus on the intuition. One way to think of the problem is that we do not know the maturity of the bond, and so we do not know what part of the curve to use: the early (today to first call) or late (call to maturity). A callable bond is somehow a combination of the curve to first call date and the curve from first call to maturity.

Exhibit 8.6, which covers February 1929, illustrates the problem. Two bonds in particular (the 4.25s of 15 October 1938/1933 and the 4.25s of 15 June 1947/1932) are trading just below or just above par. An ad hoc rule ("yield-to-worst") is to use the first call date if the current price is above par (presuming the bond will be called for certain) and the maturity date if the price is below par. Panel A shows the curve using this rule. The two bonds are trading just above par and are treated as maturing in 1938 (9.6 years, 7.9 years duration) and 1947 (18.3 years,12.8 years duration), respectively.

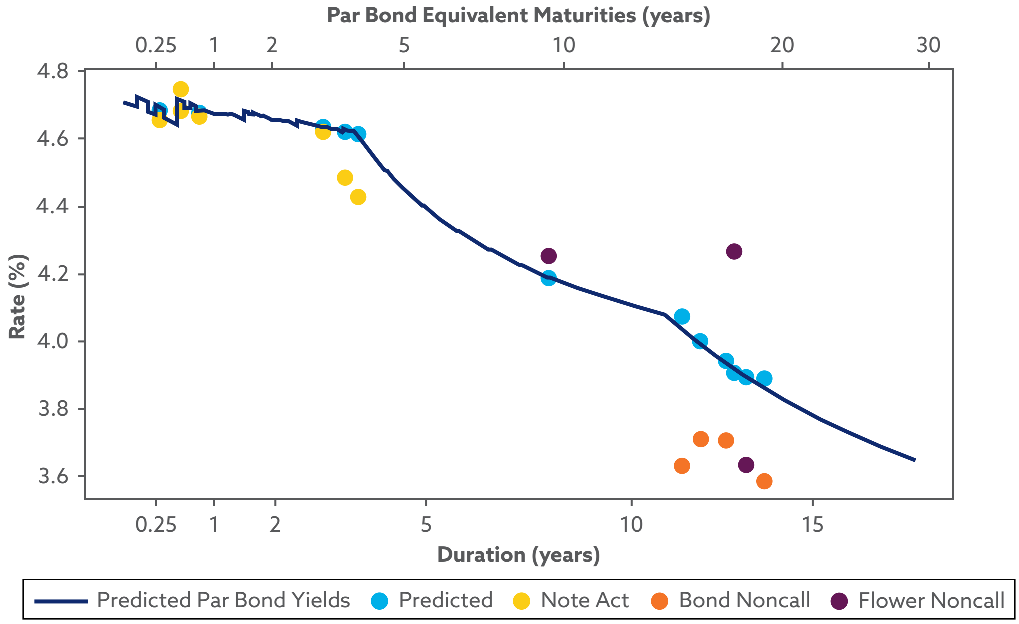
The curve in Panel A presents an issue: The 4.25s of 1947/1932 in particular are trading at a much higher yield than other bonds of similar duration and final maturity. The reason is that the 4.25s of 1947/1932 have a high option value, are trading as a combination of short (3.3 years) and long (18.3 years) maturity, and thus behave "shorter" than their final maturity. With the downward-sloping yield curve, this means they have a yield higher than noncallable bonds with that same final maturity. The other long-dated bonds, although callable, have a low probability of call and are trading more like long-dated bonds, with lower yields (given the slope of the yield curve on that date).

As a result, the yield curve is distorted, with neither the 4.25s of 1947/1932 nor the other bonds fitting well at the long end. The solution to the problem is to use the option formula of Equation 8.2, which treats the bond as a probability-weighted combination of first call and final maturity.

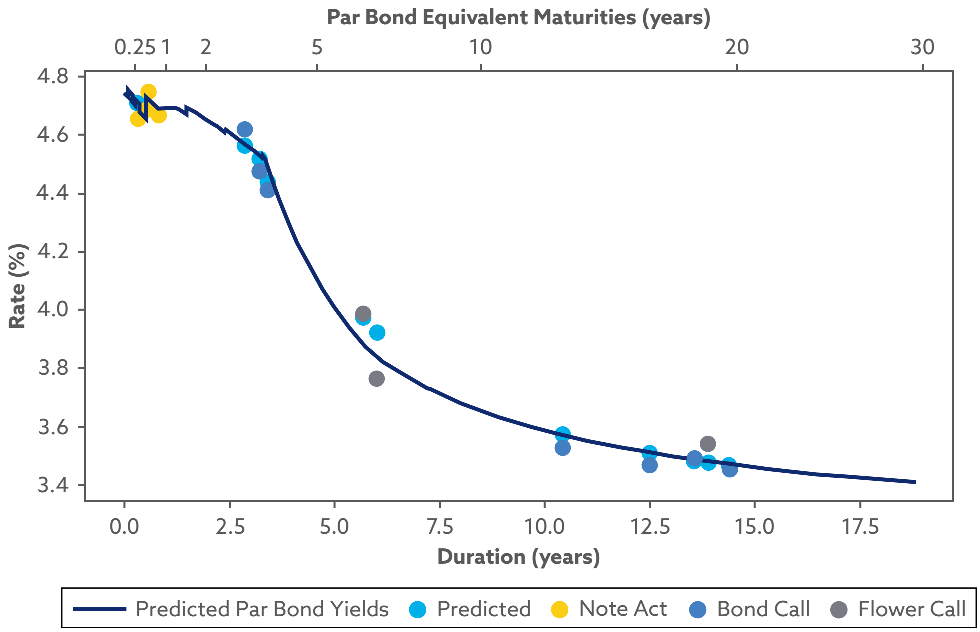
Panel B shows the results. Both bonds are a weighted average of short and long, with Macaulay durations very close to each other (5.7 and 6.0 years). We use an option-adjusted Macaulay duration to represent the length of the bonds because the bond "length" is not clearly defined. The standard definition of Macaulay duration is the present value–weighted time to repayment of cash flows and is defined only for fixed-cash-flow instruments. We can, however, use the relation between Macaulay duration and modified duration—the percentage sensitivity of PV to

Exhibit 8.6. Forward and Par Curve for 1929, Yield-to-Worst and Option-Adjusted

A. Par Bond Curve, 28 February 1929, Yield-to-Worst (%)



B. Par Bond Curve, 28 February 1929, Option-Adjusted



Source: Author calculations. CFI par bond rates and actual versus predicted yields calculated from CFI forward rates database. Forward rates estimated from CRSP Treasury bond prices using CFI methodology, as discussed in the text.

Note: For callable bonds, a normal yield option model is used to evaluate the option to first (or next) call, with an annualized level yield volatility of 0.005 (decimal) or 0.5 percentage point, which translates into roughly 3 bps per day.

changes in yield, MacDur = ModDur(1 + *y*/Freq)—to define an option-adjusted Macaulay duration for the callable bonds.

Panel B of Exhibit 8.6 graphs the yields versus option-adjusted Macaulay duration, and the bonds line up nicely. The two 4.25% bonds fit right in the middle, with duration around six years, between the shorter bonds with higher yield and longer bonds with lower yield.

Flower Bonds

The US Treasury issued bonds from 1943 to around 1968 that could be redeemed at par ($100) for the payment of estate taxes upon death of the owner. This instrument is effectively a put held by the owner of the bond—an option to put the bond back to the government at $100 upon death of the owner. These are known as "flower bonds" because of the association with flowers given at funerals. Exhibit 8.3 shows that prior to 1973, all long-dated bonds were flower bonds.

During periods of high interest rates, a normal bond with a low coupon would sell at a discount. During such a period, a flower bond would become valuable because of the effective put: If bought at a discount, it could be redeemed for $100 after the owner's death. Such a flower bond would trade at a premium relative to non-flower bonds and a premium to the discounted present value—the yield would be lower than it otherwise would be.

Not all flower bonds will trade equally as flower bonds. Consider two flower bonds, one with a lower coupon. If the flower option were removed, the two bonds would trade at (roughly) the same yield and the lower-coupon bond would have a lower price (larger discount). With the flower option, that larger discount provides a larger value to flower bond buyers, who will choose the lower-coupon and lower-priced bond over the higher-coupon bond. This higher demand will push the price of the lower-coupon bond up (and yield down) relative to the higher-coupon bond. When the lower-coupon bond is trading at a lower yield (higher price), we can conclude that the lower-coupon bond is trading with a larger flower component. In any month, we can examine all flower bonds and determine which are trading with the largest flower component—so-called flower bonds *par excellence*. Larry Fisher, the originator of the CRSP US Treasury bond database, created a list of flower bonds *par excellence*, which we used in our original Coleman et al. (1989, 1993) books and which we have updated slightly.

The lack of non-flower bonds and the rising rates during the 1960s and early 1970s makes estimating the long end of the curve challenging. The best we can do prior to 1973 is to use our list of flower bonds (given in the online appendix) to exclude flower bonds *par excellence*. This produces a curve where the long end is contaminated to the least degree possible with the flower bond effect. In 1973, the Treasury started to issue long-dated non-flower bonds (the 6.75s of 15 February 1993 issued in January 1973 and the 7s of 15 May 1998/1993 issued in May 1973). Starting in July 1973, we exclude all flower bonds and move to a non-flower curve. For returns, we create a series using flower bonds up to June 1973 and excluding flower bonds starting in July 1973 (but using the non-flower curve for June 1973 to calculate the return for July 1973).

Multiple Clienteles: 1968, Bonds, and Flowers

In the 1960s, the US Treasury curve appeared to be trading with multiple clienteles: short-end bills, medium-term notes versus bonds, and long-term flower bonds. The problem appeared to result from a 4.25% cap on bond coupons (with the Fourth Liberty Loan Act in September 1918, Congress restricted the maximum coupon for bonds but not notes), together with rising yields by the late 1960s (up to 7% for the long end by 1969, well above 4.25%). The cap effectively precluded the Treasury from issuing long-dated bonds, and between 1965 and 1971, the Treasury did not issue bonds but issued only notes with maturity up to seven years.[4] Older and lower-coupon bonds seem to have traded at a premium (lower yield) to notes with the same maturity, presumably because of scarcity of bonds. We try to manage this issue by weighting notes more heavily than bonds during this period in our nonlinear least-squares fitting. This means our curve is effectively a note curve in the intermediate part of the curve (two to seven years).

Key Takeaways

- The Coleman Fisher Ibbotson data include monthly US Treasury forward rates, continuous par bond yield curves, par bond returns of any selected maturity or duration, and other yield curves and returns as desired—for example, zero coupon, annuity, and so on—in that any "riskless" US dollar cash-flow stream can be priced over the 1926–2025 period.
- The CRSP government bond file of raw US Treasury bonds, notes, and bills is used to estimate forward interest rates with piecewise twisted forward rate functions (splines) covering the short term (seven days) to the long term (up to 32 years).
- Forward rates are used to estimate par bond yield curves over different tax regimes (partially tax exempt for 1925–1932, tax exempt for 1930–1942, and fully taxable for 1940–2025), with flower bonds (issuance 1943–1968) in which the lowest dollar prices are distorted because they could be used to pay estate taxes at par and callable bonds (issuance 1925–1984) whose maturities are option adjusted. After 1988, all US Treasury bond issued were noncallable and fully taxable, but previous issuances still existed.
- Par bond returns are calculated from the monthly continuous yield curves for selected maturities and durations, with returns consisting of the income, the modified duration times the negative yield change, and the gain or loss from rolling down the yield curve as the bonds shorten their maturity each month.

References

Black, Fischer, and Myron Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81 (3): 637–54. doi:10.1142/9789814759588_0001.

Coleman, Thomas S., Lawrence Fisher, and Roger G. Ibbotson. 1989. *US Treasury Yield Curves, 1926–1988*. Moody's Investors Service.

———. 1993. *Historical US Treasury Yield Curves*. New York: Ibbotson Associates, Moody's Investors Service.

Hagan, Patrick S., and Graeme West. 2015. "Methods for Constructing a Yield Curve." *Wilmott Magazine* (July): 70–81.

[4] In 1967, Treasury notes were redefined (by statute) to mature in one to seven years (instead of five), and in 1971, Congress authorized issuance of $10 billion of higher-coupon bonds. In 1988, Congress eliminated the 4.25% coupon limit.

CHAPTER 9. THE EQUITY RISK PREMIUM

Laurence B. Siegel
Gary P. Brinson Research Director Emeritus, CFA Institute Research Foundation

Carla S. Nunes, CFA
Managing Director and Head of Capital Markets Research, Valuation Research Corporation

Why We Care About the Equity Risk Premium

Why do we care about the equity risk premium? Why is it often described as the most important variable in finance? And why does so much effort go into estimating it, both historically and as a forecast?

The simple answer is that the equity market is the most important capital market—it is where change in the wealth of the world occurs on the margin—and the equity risk premium (ERP) is the *price of risk* in that market. Understanding the price of equity risk—that is, the additional return that people expect on the margin for assuming an additional unit of equity risk—enables us to unravel puzzles in practically every branch of finance.

Applications of the Equity Risk Premium

The ERP is the key element of the expected return on equities (the remainder of that return consisting of the riskless rate). Estimates of the expected ERP are useful in a large number of applications, some of which we will describe in further detail later in this chapter:

- asset allocation (between equity and fixed-income asset classes),
- financial planning, including projection of the long-term growth of wealth and determination of appropriate saving and spending rates,
- determination of the cost of equity capital, essential in corporate valuation settings, capital budgeting, and other aspects of corporate finance, and, overarching all of the above,
- as the crucial input to the capital asset pricing model (CAPM), which estimates the expected return on a security or portfolio conditional on knowing its beta (i.e., its systematic risk, the portion of the security's total risk that is correlated with the risk of the overall market).

Expected Return Does Triple Duty

The central place of the ERP in finance is illustrated by the "triple duty" nature of expected returns, of which the ERP is the most variable and most interesting part. In the hypothetical construct that economists call a perfect capital market, (1) the *expected return* on a risky security, (2) the *cost of capital* faced by a firm issuing that security, and (3) the *discount rate* appropriate for reducing the future cash flows from that security to their present value (or market price) are all the *same number*.

Because information about securities is imperfect and costly, as well as because of taxes, transaction costs, and differences in investors' tastes and preferences, however, the capital market is obviously not perfect—far from it! As a result, this triple equality does not hold precisely in practice. It is, however, a useful framework for linking the various aspects of finance in which the expected ERP is used.

Principal Approaches to Estimating the Expected ERP

There are four broad categories of approaches to estimating the expected ERP:

- historical,
- demand oriented,
- supply oriented, and
- surveys.

We begin with the historical approach taken by Roger Ibbotson and Rex Sinquefield (1976a, 1976b) in their classic work, the semicentennial of which we are celebrating in this book. Moreover, during much of the time since their original work, the historical model was the dominant way of estimating the ERP.

The Historical Return Approach

One of the most important reasons why Ibbotson and Sinquefield (1976a, 1976b) investigated the historical total returns on US equities starting in 1926 was simply that no one knew what the relative long-run rates of return on the principal asset classes had been. Cowles (1938) and Fisher and Lorie (1964) had collected long-run stock market returns, but they had no riskless or reference asset with which to compare them. Ibbotson and Sinquefield, by calculating return *differences*, homed in on the realized equity risk premium and other premiums discussed in Chapter 2.

Ibbotson and Sinquefield's original articles and subsequent updates (Stocks, Bonds, Bills, and Inflation®, hereafter SBBI®) revealed that the realized ERP in the United States had been quite large over the period studied, enough to justify a large allocation to stocks for investors with typical risk preferences. In fact, some of the shift in investor behavior from a fixed-income-dominant to equity-dominant investment allocation over the last half century can be attributed to Ibbotson and Sinquefield's findings.

Using Historical Returns for Forecasting

But knowing what happened in the past is of limited usefulness unless that information tells us something about the future. How is an accurate measure of the past, or realized, ERP helpful for understanding future outcomes? One possibility is that the historical record is representative of what can happen and is likely to happen in the future—in terms of risk as well as return. A statistician would call such a process "stationary."

In their early works, Ibbotson and Sinquefield proposed a forecasting method based on the premise that the ERP is stationary. We call this approach the historical method or historical model. To justify it, Ibbotson and Sinquefield argued that past realizations of the ERP are a good guide to the future because of the following:

- There is no compelling evidence that either the amount of risk in the market or the price of risk (defined earlier) has changed systematically over time or would be expected to change materially in the future.
- Over time, investors conform their expectations to what is proven to be realizable in financial markets, so historical returns are at least a plausible guide to investors' expectations.[1]
- A history as long as 1926–1974 (initially) or, better yet, 1926–2025 (today) contains most of the event types and economic conditions that investors should expect in the future. These types of events, including bull and bear markets, crashes, periods of stability, high and low inflation, wars, recessions, financial crises, and so forth, have been reflected in market returns that are averaged together to form the historical return data. Thus, the historical or realized ERP is a good guide to what investors can expect in the future.

At the time of Ibbotson and Sinquefield's original work, other means of estimating the expected (future) ERP, such as the dividend discount model, had produced forecasts that were way too low. A better method was needed.

Although many investors now expect that the US ERP will be lower in the future than in the past century, the past provides a helpful framework and baseline for developing future expectations.

Applying the Historical Model

Applying the historical, or what Ibbotson and Sinquefield called the "future equals past," model is conceptually simple. The expected value, or mean, of the ERP in the future is equal to the average ERP realized in the past. If you simply want a point estimate of the expected ERP, then according to this model, you are done. But risk, as well as return, must be taken into consideration in any forecast. Ibbotson and Sinquefield's 1976 articles and SBBI® assume that the *probability distribution*, as well as the mean, of the realized ERP will be the same in the future as it was in the past.

Adjustments can be made to the historical model. An example is in Chapter 20, wherein the US historical returns are modified by making an international adjustment because the United States was the best-performing country rather than a typical country. Another adjustment might be to take out the increase in the price-to-earnings ratio (P/E) that occurred in the United States in the past century, and we make this adjustment in what we call the adjusted historical model, which we cover later in this chapter, in the section on the supply of returns.

[1] See Ilmanen and Maloney (2026) for a thorough treatment of how investors form return expectations.

The Supply and Demand for Returns: An Introduction

We now provide a brief description of the supply and demand framework introduced by Ibbotson, Diermeier, and Siegel (1984) and Diermeier, Ibbotson, and Siegel (1984). This background is needed to understand the rest of this chapter.

In these two articles, the authors introduced the concept of the demand for returns (what return investors require in order to take the risks inherent in a given security) and the supply of returns (the cash flows that corporations provide to their stockholders and other capital providers). Although the traditional finance literature refers to the supply and demand for *capital*, we flip this concept around, both in this discussion and in a rich vein of literature that developed in the wake of the 1984 articles: We regard buyers of securities as demanders of capital market *returns* and sellers or issuers as suppliers of those returns. Analyzing the supply and demand for returns in this way sheds greater light on the questions we are trying to answer in this book.

The Demand for Returns

The concept of an equity risk premium arises from the demand side of this equation. If investors are risk averse, they require a higher return from investing in risky assets than from holding safe ("riskless") ones. That difference is the ERP if the risky asset under consideration is an equity index and the riskless asset is a Treasury bill, note, or bond.[2]

Capital Asset Pricing Model

The CAPM, developed by William Sharpe (1964) and several other researchers at about the same time,[3] formalizes this relationship. The CAPM says that the expected return on a risky security is equal to the riskless rate plus the security's beta (a measure of its market risk, the part of its total risk that is correlated with movements of the equity market index) multiplied by the ERP:

$$E(r_p) = r_f + \beta_p(\text{ERP}), \tag{9.1}$$

where

$E(r_p)$ is the expected return on security or portfolio p

β_p is the beta of security or portfolio p

r_f is the risk-free rate

ERP is given by $E(r_m) - r_f$—that is, the expected return on the market index minus the riskless rate

---

[2] This chapter refers often to US Treasury securities or government bonds. The same concept would apply, however, if we were discussing these topics for other countries. Generally speaking, we are referring to government securities issued by countries whose sovereign debt is largely regarded as free of default (or riskless).

[3] The CAPM was developed in the early 1960s by Sharpe (1964), Treynor (1962), Lintner (1965a, 1965b), and Mossin (1966). For a good historical perspective on the development of the CAPM, see Perold (2004).

The CAPM does not specify how the ERP should be estimated; instead, it assumes one can obtain the needed estimate from a separate, independent source. The challenge posed by the CAPM's need for an ERP input is the original reason why researchers sought to obtain an accurate estimate of the (future or expected) ERP. Even though the CAPM is no longer the only equilibrium model of asset returns that practitioners use, it remains centrally important to the problem of estimating the expected ERP. By recognizing that investors require an additional return for investing in risky as opposed to riskless assets, the CAPM sets up the question of *why* a (nonzero) ERP exists, as well as what the right number is.[4]

Popularity Asset Pricing Model

Another demand-side approach, merging ideas in the CAPM and in Ibbotson et al. (1984), is the popularity asset pricing model (PAPM) of Ibbotson, Idzorek, Kaplan, and Xiong (2018). Covered in Chapter 18 of this book, the PAPM merits a brief summary here. In the PAPM, the demand for an asset reflects asset characteristics to which a given investor is attracted or by which they are repelled, including not only risk and return but also such characteristics as liquidity (or lack thereof), tax treatment, reputation of the company, and ethical or environmental considerations. Aggregating across investors, the more popular or well liked the asset is, for any of these reasons, the *lower* the expected return (because the asset's popularity has caused its price to be bid up to a level that reduces future returns).

Mehra–Prescott or Pure Utility Theory Model

A demand-side approach exists that does result in a numerical estimate of the expected ERP, although it is rarely used in practice. The work of Mehra and Prescott (1985) looks outside the financial markets to utility models with reasonable risk-aversion parameters to ascertain what investors would rationally demand, or require, in order to take on the risk of equities instead of riskless assets. These authors based their method on the observed price of risk in macroeconomic settings, but their ERP estimates are an order of magnitude lower than those obtained through the other methods described here, so we do not pursue this method any further.

The Supply of Returns

Supply models of the expected ERP measure what corporations offer to would-be investors in terms of expected cash flows (distributable profits). The classic dividend discount model (DDM) of Graham and Dodd (1934) is an example of a supply model because it takes as an input the cash flows to the investor from the corporation. Ibbotson and his coauthors have contributed to enhancements and modifications to the DDM (see, e.g., Diermeier et al. 1984; Ibbotson and Chen 2003; Straehl and Ibbotson 2017), thus adding to the supply-model literature. This subsection discusses these works, along with contributions from other authors that are supportive of this general approach.

Supply-based models of the ERP have become increasingly influential in the last generation and have emerged as a complement to Ibbotson and Sinquefield's historical approach. In general,

[4] Note that the CAPM is a single-period model, generally requiring a short-term estimate of the ERP (one that is derived by subtracting the yield on a short-term reference or "riskless" asset). This chapter focuses on long-term estimates of the ERP.

supply models have in common the characteristic that they use information from the real economy, rather than market prices as is the case with purely historical models.

The intense interest in supply models in the late twentieth and early twenty-first centuries emerged when many analysts noticed that the historical model was producing overly high ERP estimates. The work of Ibbotson and Chen (2003), described later, is an example, but movement in the direction of supply models began earlier, with Campbell and Shiller (1988), Fama and French (1988), and Jeremy Siegel (1994). In the current century, the works of Shiller (2000), Asness (2000, 2003), Arnott and Bernstein (2002), Cochrane (2011), and Ilmanen (2011) stand out as landmark efforts. Support for supply models continued to increase in the wake of the market collapses of 2000–2003 and 2007–2009.[5]

Diermeier, Ibbotson, and Siegel (1984): A Quick and Dirty Supply Model

The first and most basic supply model is that of Diermeier et al. (1984). The authors started with the observation that equity total returns must equal dividend yields plus the growth rate of dividends. This latter quantity cannot in the infinitely long run exceed the growth rate of the overall economy or else dividends would crowd out all other economic claims. Diermeier et al. completed the model, which we now regard as a "quick and dirty" estimate, by asserting that the expected equity total return equals the dividend yield minus new issues net of share buybacks plus the GDP growth rate as a proxy for dividend growth. The ERP is the expected equity total return minus the riskless rate.

Grinold and Kroner (2002) and Grinold, Kroner, and Siegel (2011) embellished this model, adding new data on buybacks in light of their dramatically increased importance in recent decades.[6]

Ibbotson and Chen (2003): The Adjusted Historical Model

One potential flaw of the historical model is that it implicitly projects forward the very substantial past increase in the valuation ratio (P/E) of the market over 1926–2025 indefinitely into the future. Because such a further rise is unlikely to happen again starting at today's high valuations, one can adopt an *adjusted historical model* instead: Take the historical results and subtract that part of the return attributable to the rise in P/E.

As of this writing, the historical 1926–2025 arithmetic mean ERP, based on the S&P 500 and predecessor indices and using long-term bonds as the reference asset, was 6.4%. Subtracting the arithmetic mean annual rate of P/E expansion, which was 1.1%, we arrive at an expected arithmetic mean ERP of 5.3%.[7] This is the current result using the adjusted historical method.

The adjusted historical method can be regarded as a type of historical model or as a supply-related model. Ibbotson and Chen (2003) showed that at least with respect to past returns,

---

[5] Laurence Siegel (2017, pp. 9–12) provided more detail on this movement in a section called "Time-Varying Premia and the DDM Counterrevolution." The supply models we have discussed, as well as those covered in Siegel (2017), have become prominent—even dominant—in ERP estimation as practiced by investment managers and in corporate finance and valuation settings in the twenty-first century.

[6] Chen and Obizhaeva (2022) provided a good literature review on the rising importance of buybacks.

[7] The 1.1% estimate of annual P/E expansion was calculated by Ibbotson and Chen (2003) as the compound annual growth rate of Shiller's CAPE (see footnote 10) over 1926–2025. This estimate differs from the one shown in the "Implementation Illustration" section, which is based on three-year average S&P 500 earnings.

the adjusted historical model is consistent with a supply approach. They outlined six different ways of decomposing historical equity returns. One of those ways (which they called the earnings model) decomposes the historical equity return into four components: income return, inflation, growth in real earnings per share (EPS), and growth in the P/E. The adjusted historical ERP sums the first three components but leaves out the growth in P/E.

Ibbotson and Chen noted that the historical capital gain return of the stock market equals EPS growth plus P/E expansion—so if we take out P/E expansion, we are looking only at information from corporations (the real economy). A supply-side input is thus inherent in the adjusted historical model.[8]

Reflecting the influence of the supply side, the output of the adjusted historical model has been labeled the "supply-side ERP" in various Ibbotson Associates and Morningstar publications and has been referred to as such in court cases and other legal and regulatory contexts.[9] It is the method used in the "Implementation Illustration" section of this chapter.

Straehl and Ibbotson (2017): Adding in Buybacks as a Source of Return

In a follow-on article, Straehl and Ibbotson (2017) enhanced the supply literature by showing that over a longer time period (1871–2014), total cash payouts to the investor—dividends plus share buybacks—are the key drivers of long-run equity returns. They showed that "total payouts per share (adjusted for the … decrease [in the number of shares resulting from] buybacks) grew in line with economic productivity, whereas aggregate total payouts grew in line with GDP" (Straehl and Ibbotson 2017, p. 32). This analysis bolsters the contention that information from the real economy is useful and additive in predicting future stock returns.

Straehl and Ibbotson (2017) also introduced a new valuation ratio, cyclically adjusted total yield (CATY). This measure adds buybacks to dividend payments to arrive at an improved measure of cash flow to the investor, which they called *total yield*. The authors showed that CATY "predicts

---

[8] The remainder of the historical total return is the dividend yield, which is also partly sourced in the real economy. The dividends themselves come from corporations, but the dividend *yield* involves dividing by a market price, so the adjusted historical model is only partly a supply model. Stated another way, the "supply-side ERP" is a historical method with an adjustment based on the idea that the P/E expansion over the last century is unlikely to be repeated, and it reconciles to the supply-side model outlined in Ibbotson and Chen (2003) when applied historically. When applied to the future, these two models can diverge because the earnings or dividend growth estimates may come from a different source than historical stock returns.

[9] Specifically, the adjusted historical estimate by Ibbotson Associates and Morningstar consists of the "historical equity risk premium minus [the change in] price-to-earnings ratio calculated using three-year average earnings" (Ibbotson Associates 2005, p. 258). This estimate has emerged as one of the most frequently used ERP estimates in US Delaware Court of Chancery appraisal cases.

The Delaware Court of Chancery is often the forum of shareholder disputes in the United States, with valuation disagreements at the core of many of these cases. Accordingly, the business valuation and appraisal communities and investment advisers engaged in pricing mergers and acquisitions pay close attention to the inputs used in those valuations, including cost-of-capital-related decisions. The Delaware Court of Chancery has expressed in some cases a preference for using a supply-side ERP that is identical to the adjusted historical method described in this chapter. The first time the court decided in favor of the supply-side ERP (rejecting the historical ERP also published in the SBBI) was in April 2010 in the case Global GT LP v. Golden Telecom, Inc., 993 A.2d 497 (Del. Ch. 2010; https://cases.justia.com/delaware/court-of-chancery/137130-0.pdf?ts=1462311681). A sequence of court decisions that adopted the same approach followed. Nevertheless, the parties of any case are still advised to support their assumptions, and the "supply-side ERP" should not be automatically presumed to be the accepted method by the court.

changes in expected returns at least as well as the … CAPE" (Straehl and Ibbotson 2017, p. 32).[10] Given the decreased importance of dividends and greater prevalence of buybacks in recent years, we expect CATY to outperform CAPE (cyclically adjusted P/E) even more as a forecaster of future returns than it has in the past.[11]

Implied ERP

Yet another supply-based model of the expected ERP was set forth by Damodaran (2008), whose "implied ERP" is conceptually based on the DDM. It begins by solving for an implied cost of equity capital (i.e., expected total return on equity), which is the sum of the riskless rate and the ERP. This implied cost of equity capital is the discount rate that equates the present value of consensus expected earnings (or dividends and buybacks) for a stock market index with the current level of that index. The implied ERP is then obtained by subtracting the current yield on the riskless asset.[12]

Surveys of the Expected ERP

Setting aside economic modeling, one way to find out what people expect from the markets is to ask them. Thus, surveys have become an alternative or supplement to the historical, supply, and demand approaches.

Surveys of expected returns have been around for a long time, and one in particular has gained traction among certain communities of investors, corporations, and advisers. Since 2009, Pablo Fernandez and his coauthors have conducted an annual survey of finance and economics professors, which was expanded to include analysts and corporate managers in 2010. In these surveys, respondents are asked what ERP they are using to calculate the *required* return to equity in different countries at that point in time. Since 2013, with the exception of 2016, the annual survey has also asked for the riskless rate assumption being used along with the reported ERP.

Fernandez (2004, 2023) portrays the ERP as embracing four different concepts: historical, expected, required, and implied ERP. In the context of the CAPM, he says, the model assumes that the required ERP is the same as the expected ERP.

The survey has gained some acceptance as a benchmark for ERPs, given its breadth of coverage, particularly outside the United States and among some US multinational corporations and private equity investors that need to value foreign operations. For example, the 2026 survey (Fernandez, Habibian, and Acín 2026) reported riskless rates and ERP results for 97 countries.

---

[10] CAPE was developed by Robert Shiller at Yale University. It divides the current price on the S&P 500 by the trailing average of 10 years of inflation-adjusted earnings, smoothing out economic cycles.

[11] An unpublished paper (Siegel 2018), with a funny title ("She Caught the CATY and Left Me a Mule to Ride," a riff on a lyric from a Blues Brothers song), provides more color on CATY and adds a third source of cash flow to the investor: cash takeovers of companies in the index by companies (including private equity funds) not in the index. In Siegel's interpretation, this last source of cash boosts the supply-side estimate of the ERP by a not insignificant 0.6%.

[12] In his application of the DDM model, Damodaran assumes that the long-term growth rate for S&P 500 earnings (the perpetuity growth rate after the fifth year) is equal to the 10-year US Treasury yield. The basis for such an assumption is weak and during times of market dislocations may result in distorted ERP estimates.

Unfortunately, these surveys do not make clear whether the ERP forecast being solicited is of the arithmetic or geometric mean or relative to bonds or bills. The absence of this information makes the survey results difficult to interpret.

Issues in Numerical Calculation of the ERP

Measuring the ERP—the (past or prospective) difference between the returns on two assets or asset classes—sounds simple, but it is not. In this section, we examine some of the issues involved in calculating the ERP.

Arithmetic vs. Geometric Mean Differences (Risk Premiums)

One of the most contentious issues is whether the *arithmetic mean difference* or the *geometric mean difference* is the relevant measure for the ERP. As mentioned in Chapter 2, both are relevant but for different applications. Before exploring the calculation of *differences* between asset return series, let us first recall from Chapter 3 the distinction between the arithmetic and the geometric average (mean) rate of return on a *single series*. The arithmetic mean is the simple average of the periodic (say, annual) returns on an asset; the geometric mean is the compound rate of return that equates the beginning and ending values of an investment in that asset.

Chapter 2 reported that the Ibbotson Large Cap stocks total return index had an arithmetic mean return of 11.8% and a geometric mean return of 10.1% over the 1926–2025 period.[13] Over the same period, 20-year US Treasury bonds, which have less risk (a lower standard deviation), had an arithmetic mean total return of 5.4% and a yield return of 4.9%. The corresponding geometric mean total and yield returns were both 4.9%. The long-horizon arithmetic average ERP over this full-century period was thus 11.8% − 4.9% = 6.9%—that is, the arithmetic mean stock total return minus the arithmetic mean bond yield return.

When two arithmetic means are being compared, only the arithmetic difference (for example, $R_{A,Stocks} - R_{A,Bond\ Yield}$) is relevant. That calculation is how we arrived at the 6.9% arithmetic mean estimate for the ERP relative to bonds. To compare two geometric means, however, one could choose either the arithmetic difference between them, $R_{G,Stocks} - R_{G,Bonds}$, or the geometric difference, $\frac{1+R_{G,Stocks}}{1+R_{G,Bonds}}-1$. The geometric difference provides a more precise answer and is always smaller than the arithmetic difference (for time series that have a nonzero variance, which is the case for asset returns).[14] **Exhibit 9.1** shows the ERP and other risk premiums.[15]

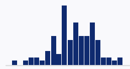
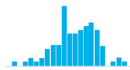
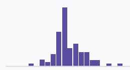
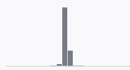
---

[13] These are summary statistics of annual returns, not annualized summary statistics of monthly returns.

[14] The geometric mean approaches the arithmetic mean when there is little variability in the time series.

[15] The arithmetic means shown are the mean of the risk premium time series created by subtracting the two basic asset class annual returns, and the geometric means shown are the geometric mean of the risk premium time series created as the geometric difference of the two basic asset class annual returns—that is, $\frac{1+r_A}{1+r_B}-1$, where r_A and r_B are the annual returns on asset classes A and B, respectively. The population standard deviations shown are of the latter time series (the one created by geometric differencing).

Exhibit 9.1. Realized Risk Premiums on Ibbotson Equity and Bond Indices and Real Interest Rates, 1926–2025

	Compound Annual Return	Arithmetic Mean Annual Return	Risk (standard deviation)	
Long-horizon equity RP	4.9%	6.9%	18.4%	
Short-horizon equity RP	6.6%	8.5%	18.7%	
Small-stock RP	0.8%	2.6%	13.1%	
Bond-horizon RP	1.6%	1.6%	1.3%	
Real interest rate	0.3%	0.3%	3.9%	
				–90 0 90

Source: Calculated by the author from Ibbotson Indices data.

Notes: RP = risk premium. The risk measure is population, not sample, standard deviation.

Bonds vs. Bills as the "Riskless" Asset

In Chapters 2 and 4, as well as elsewhere in this book, we chose to use the 20-year US Treasury bond as the "riskless" or reference asset to which risky assets can be compared. The choice of a riskless asset should depend on the time horizon of the person doing the choosing. No asset is completely riskless (the prices of long-term bonds fluctuate; the yields of short-term bills fluctuate).[16] When one is estimating the ERP, however, the default-free bond should match as closely as possible the time horizon of the investor.

Because we are concerned with the picture facing long-term investors, we generally use the 20-year US Treasury bond as a proxy for riskless investing in this book. The maturity of such a bond may reasonably match (as closely as we can with a single bond maturity) a long-term investor's time horizon. In the context of corporate finance and valuation, where businesses are often valued as going concerns, the ERP is also typically estimated relative to a long-term government bond as the proxy for the riskless asset, as discussed later in this chapter.

[16] Treasury Inflation-Protected Securities (TIPS) are sometimes described as truly riskless, but they are not riskless. Although they are indexed 1:1 to the Consumer Price Index for All Urban Consumers (CPI-U), TIPS have risks, as enumerated by Pfau (2011). Importantly, liquidity issues in the TIPS market are well documented, with various authors developing models to estimate the TIPS liquidity premium for the US and other markets. See, for example, Andreasen, Christensen, and Riddell (2021). In addition to the risks listed by these authors, the main measure of inflation used by the US Bureau of Labor of Statistics, the CPI-U, may not match the inflation rate experienced by any given investor.

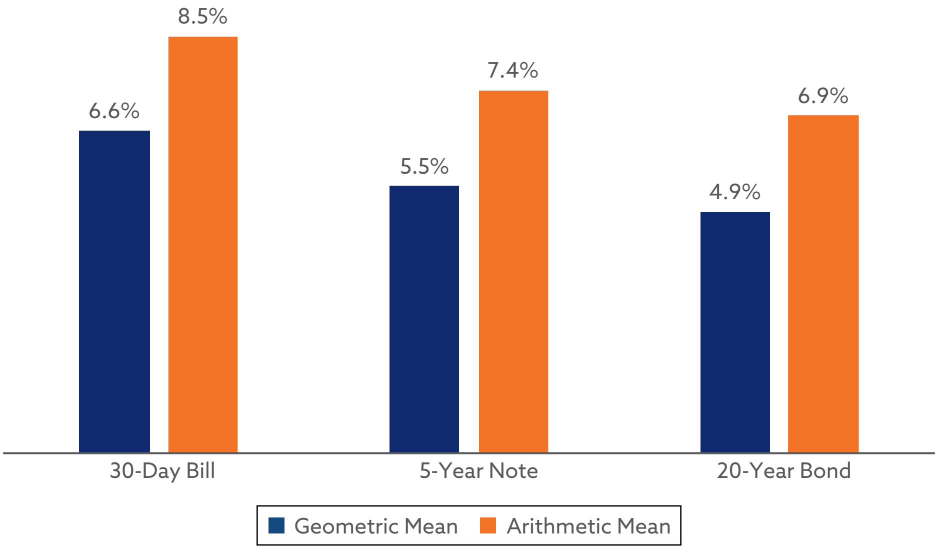
Thus, our entire discussion of the ERP in this book is based on the choice of a 20-year bond as the "riskless" asset. Some investors have other time horizons (or no particular horizon), however, and an ERP calculated on the basis of a different bond maturity can be useful for certain applications.[17]

In **Exhibit 9.2**, we illustrate the concept that realized ERPs will differ depending on the maturity selected for the riskless asset. This important concept is often ignored in practice but can result in significantly different ERP estimates.

Use of Bond Yield vs. Bond Total Return in Calculating the ERP

As was done in SBBI (for example, Ibbotson and Harrington 2020), we use the *yield* on the bond at the time of "purchase," not the total return over the time period the bond is held, as the relevant number to subtract from the equity total return in calculating the ERP. We do that because the yield at time of purchase is the expected return on the bond; any subsequent gains or losses from changes in market yields are *unexpected*. Because we are using measures of the realized ERP to ascertain what investors should and do expect in the future—that is, to make forecasts—we define the realized ERP as the actual return on stocks minus the return that the

Exhibit 9.2. Historical ERP: Geometric and Arithmetic Mean Premiums in Excess of US Treasury Bill, Note, and Bond Maturities, 1926–2025



Sources: Data from Ibbotson Large Cap stocks total return index and 30 Day Treasury Bills, CFI 5 Year T Notes, and CFI 20 Year T Bonds yield indices.

[17] For example, shareholder disputes involving the computation of lost profits or business damages (e.g., for patent or trademark infringement) tend to be focused on a shorter period (i.e., the period when losses/damages were incurred, which is often less than 20 years). In such cases, when quantifying the (present) value of such losses/damages, advisers will often use a maturity on the riskless asset commensurate with the period being evaluated.

investor *expected* on a particular long-term bond (i.e., the income return component, a reasonable equivalent for the expected yield at a given time).

In other words, because we are trying to ascertain what investors expected, we do not subtract the total return that the investor actually earned on the bond because that result includes unexpected surprises. Instead, we subtract the total return that the investor rationally expected at the beginning of the period.[18]

What Historical Period Is Relevant?

As we saw in Chapter 4, the decision of what historical period to use for estimating the ERP involves a tradeoff. The longer the period, the "tighter" the estimate (that is, the standard error of the estimate declines with the square root of the number of years). The farther back one goes in time, however, the more likely the estimate will include data from periods that are irrelevant because they are so long ago.

When the first Ibbotson and Sinquefield studies were published, 1926 to the present (at the time) was already a fairly long period, so investors and others in search of long-term historical data usually used that whole period.[19] There are reasons one might use a shorter or longer period, however. Starting the period after the chaos of the Great Depression and World War II can be defended on the basis of relevance. **Exhibit 9.3** shows how the ERP estimates change as one shortens the period. The average returns fluctuate, but the standard error goes up monotonically as the period becomes shorter.

A longer period has been used to support the claim of stationarity—that the relative returns of equities and bonds, unlike the absolute returns, are close to constant (Siegel 1994). Using more recently collected data, however, McQuarrie (2024) challenged this claim, showing that the realized ERP was lower in the nineteenth century relative to more recent periods (he reiterates this assertion in Chapter 10 of this book). If one links McQuarrie's data with the Ibbotson Indices, the resulting ERP over the whole 233-year period is surprisingly low.

In selecting a historical period, investors should balance the need for relevant data from markets that resemble those of today against the desire to use an ERP that is estimated with a low standard error.

---

[18] Not everyone uses this approach to calculate realized ERPs. Some academics instead subtract the total returns on bonds from the realized equity returns when computing realized ERPs. See, for example, Dimson, Marsh, and Staunton (2025).

[19] Prior to September 2008, it was quite common for valuation practitioners to be comfortable using a realized ERP estimate (based on data starting in 1926) that was updated once each year. Following the bankruptcy of Lehman Brothers in September 2008, however, valuation analysts and other practitioners began to question the stability of their ERP estimate. In the aftermath of the 2008 global financial crisis (GFC), other sources of ERP estimates began being used in conjunction with or in lieu of the realized ERP with 1926 as the starting point. Some practitioners continued to use the historical series but began using different period averages as a proxy for the future. These include, for example, long-term averages starting after World War II, after 1957 (when the S&P 500 was created), or after 1963 (following the creation of the Compustat database).

On the institutional investor side, overly high or unstable ERP estimates were typically moderated by blending in supply-side estimates, and this practice began long before the GFC. Some investment bankers use shorter historical periods to calculate realized ERPs.

Exhibit 9.3. Historical (Geometric vs. Arithmetic) ERP in Excess of 20-Year Treasury Yield, Calculated over Various Periods

	Number of Years	Geometric Mean	Arithmetic Mean	Standard Error of Estimate
1926–2025	100	4.9%	6.9%	1.8%
1950–2025	76	5.4%	7.1%	2.1%
1960–2025	66	4.0%	5.3%	2.0%
1970–2025	56	4.3%	6.0%	2.2%
1980–2025	46	5.7%	7.5%	2.4%
1990–2025	36	5.8%	7.7%	2.8%
2000–2025	26	4.0%	5.8%	3.5%

Source: Data from Ibbotson Large Cap stock total return index and CFI 30 Day Treasury Bills, CFI 5 Year T Notes, and CFI 20 Year T Bonds yield indices.

Implementation Illustration

As discussed previously, the Ibbotson and Chen (2003) adjusted historical ERP (removing the growth in P/Es) became known as the "supply-side ERP" in various Ibbotson Associates and Morningstar publications, as well as subsequent publications by Duff & Phelps (later rebranded as Kroll).[20] To this day, practitioners refer to this method in court cases and other legal and regulatory contexts. For the purposes of this section, we will also call this method the "supply-side ERP."

To estimate the geometric mean expected ERP using the adjusted historical model or supply-side ERP, one would

- decompose historical equity returns into income return, inflation, EPS growth, and P/E expansion;
- subtract (geometrically) the rate of P/E expansion; and
- subtract (geometrically) the income return on 20-year US Treasuries.

To estimate the rate of P/E expansion, Kroll, following the example of Ibbotson Associates and Morningstar, uses a three-year average of S&P 500 earnings in the denominator rather than a single year's earnings. This practice smooths earnings volatility resulting from nonrecurring

[20] Morningstar published this estimate from 2006 (after acquiring Ibbotson Associates) until 2013, at which point it discontinued its cost-of-capital-related publications. Duff & Phelps, LLC, a global financial advisory firm, took over the work of updating those annual estimates in 2014. Following the acquisition of Kroll, LLC, in 2018, Duff & Phelps decided to rebrand itself as Kroll, a process initiated in 2021 and completed in 2022. To avoid ambiguity, Duff & Phelps, LLC, should not be confused with the unrelated Duff & Phelps Investment Management, a wholly owned subsidiary of Virtus Investment Partners.

items and changes in profitability and allows the earnings number used in the P/E to better reflect a normalized (long-term) trend. Thus, we have the following:

- Numerator (P): the current level of the S&P 500 index (or an equivalent index)
- Denominator (E): three-year average of index earnings based on (1) last year's reported earnings on the S&P 500, (2) the forecast of the current year's reported S&P 500 earnings, and (3) one-year forward projection of reported S&P 500 earnings

Using this methodology, Kroll's estimate of the compound annual growth rate of the P/E of the S&P 500 (and predecessor indices) from 1926 to 2025 is 0.85% per year.[21] This P/E growth rate can then be subtracted (geometrically) from historical S&P 500 total returns to arrive at an adjusted historical equity return:

$$\text{AHR} = \frac{(1 + \text{TR}_{S\&P\,500})}{(1 + g_{P/E})} - 1, \tag{9.2}$$

where

AHR = adjusted historical estimate of equity returns ("adjusted historical return")

$(1 + \text{TR}_{S\&P\,500})$ = geometric average return of equity index

$g_{P/E}$ = growth rate of P/E of equity index

Using this P/E growth rate of 0.85% and the 10.1% geometric mean return (rounded) on the Ibbotson Large Cap stock total return index for 1926–2025, we estimate an adjusted historical return (which we use as a proxy for the supply-side equity expected return) of $\frac{1 + 10.1\%}{1 + 0.85\%} = 9.2\%$. Then, subtracting (geometrically) the long-term bond yield, we arrive at an estimate of the geometric mean "supply-side" ERP:

$$\text{ERP}_{SS} = \frac{(1 + \text{AHR})}{1 + Y} - 1, \tag{9.3}$$

where

ERP_{SS} = geometric "supply-side" ERP

AHR = adjusted historical return

Y = 20-year Treasury bond yield[22]

In this illustrative example, the resulting geometric average adjusted historical return (or as it is known in practice, the "supply-side" ERP) is

$$\frac{(1 + 9.2\%)}{(1 + 4.9\%)} - 1 = 4.1\%. \tag{9.4}$$

[21] This estimate is from the Kroll Cost of Capital Navigator. Kroll drew historical S&P 500 earnings from two sources: (1) 1925–87 earnings estimated based on Wilson and Jones (2002) data and (2) 1988–2025 reported S&P 500 earnings from S&P Global.

[22] Yield as of 31 December 2025 (rounded).

Summary

We have presented a variety of ways to estimate the equity risk premium, including pure historical, adjusted historical, and other demand- and supply-oriented approaches. The results do not differ radically among the approaches, suggesting that all of them have something to contribute to the dialogue about how to best make this calculation. Most practitioners, in both investment management and corporate finance, use more than one of the methods presented here, including historical and forward-looking estimates and combinations thereof.

Key Takeaways

- The equity risk premium is the key element of the expected return on equities. Estimates of the expected ERP are useful in a large number of applications. These include asset allocation, financial planning, an element of the discount rate used in valuation, determination of the cost of equity capital (for corporate finance purposes), and as the crucial input to the CAPM.
- The expected or future ERP is a forward-looking concept, and no single universally accepted methodology exists for estimating it. It may or may not be equal to the realized or historical ERP.
- There are four broad categories of approaches to estimating the expected ERP: historical, demand oriented, supply oriented, and surveys.
- Most practitioners, in both investment management and corporate finance, use a combination of these methods, including both historical and forward-looking estimates.
- Realized ERPs will differ depending on how they are being measured. Differences arise from (1) the choice of a time horizon, that is, whether the riskless rate is considered to be the yield on a long bond, short-term Treasury bill, that is, some other maturity; (2) the use of geometric versus arithmetic averages; (3) the historical look-back period selected; and (4) the use of total return versus yield return on US Treasuries.
- Ibbotson and Sinquefield's (1976a, 1976b) original work was seminal in establishing a methodology for estimating the ERP and other risk premiums. Over much of the time since their original work, their ERP estimate, using the historical model, was dominant. Although other methods are now used in combination with the original one, their groundbreaking research is the foundation for capital market decisions still being made today.

References

Andreasen, Martin M., Jens H. E. Christensen, and Simon Riddell. 2021. "The TIPS Liquidity Premium." *Review of Finance* 25 (6): 1639–75. doi:10.1093/rof/rfab018.

Arnott, Robert D., and Peter L. Bernstein. 2002. "What Risk Premium Is 'Normal'?" *Financial Analysts Journal* 58 (2): 64–85. doi:10.2469/faj.v58.n2.2524.

Asness, Clifford S. 2000. "Stocks versus Bonds: Explaining the Equity Risk Premium." *Financial Analysts Journal* 56 (2): 96–113. doi:10.2469/faj.v56.n2.2347.

———. 2003. "Fight the Fed Model." *Journal of Portfolio Management* 30 (1): 11–24. doi:10.3905/jpm.2003.319916.

Campbell, John Y., and Robert J. Shiller. 1988. "Stock Prices, Earnings, and Expected Dividends." *Journal of Finance* 43 (3): 661–76. doi:10.1111/j.1540-6261.1988.tb04598.x.

Chen, Alvin, and Olga A. Obizhaeva. 2022. *Stock Buyback Motivations and Consequences: A Literature Review*. Charlottesville, VA: CFA Institute Research Foundation. doi:10.56227/22.1.6.

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

Cowles, Alfred C., III. 1938. *Common-Stock Indexes, 1871–1937*. Bloomington, IN: Principia Press.

Damodaran, Aswath. 2008. "Equity Risk Premiums (ERP): Determinants, Estimation and Implications, September 2008 (with an October Update Reflecting the Market Crisis)." Working paper, NYU Stern School of Business. https://pages.stern.nyu.edu/~adamodar/pdfiles/papers/ERPfull.pdf.

Diermeier, Jeffrey J., Roger G. Ibbotson, and Laurence B. Siegel. 1984. "The Supply of Capital Market Returns." *Financial Analysts Journal* 40 (2): 74–80. doi:10.2469/faj.v40.n2.74.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2025. *Global Investment Returns Yearbook 2025*. Zürich: UBS.

Fama, Eugene F., and Kenneth R. French. 1988. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics* 22 (1): 3–25. doi:10.1016/0304-405X(88)90020-7.

Fernandez, Pablo. 2004. "Market Risk Premium: Required, Historical and Expected." IESE Business School (6 October). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=601761.

———. 2023. "Equity Premium: Historical, Expected, Required and Implied." Working paper. doi:10.2139/ssrn.933070.

Fernandez, Pablo, Amir Habibian, and Antonio Fdez. Acín. 2026. "Survey: Market Risk Premium and Risk-Free Rate Used for 97 Countries in 2026." SSRN (13 April). https://ssrn.com/abstract=6570342.

Fisher, Lawrence, and James H. Lorie. 1964. "Rates of Return on Investments in Common Stocks." *Journal of Business* 37 (1): 1–21. doi:10.1086/294647.

Graham, Benjamin F., and David F. Dodd. 1934. *Security Analysis*. New York: McGraw-Hill.

Grinold, Richard, and Kenneth Kroner. 2002. "The Equity Risk Premium: Analyzing the Long-Run Prospects for the Stock Market." *Investment Insights* 5 (3): 1–24. Barclays Global Investors.

Grinold, Richard C., Kenneth F. Kroner, and Laurence B. Siegel. 2011. "A Supply Model of the Equity Premium." In *Rethinking the Equity Risk Premium*, edited by Brett Hammond, Martin Leibowitz, and Laurence Siegel. Charlottesville, VA: CFA Institute Research Foundation.

Ibbotson Associates. 2005. *Stocks, Bonds, Bills, and Inflation: Valuation Edition 2005 Yearbook*. Chicago: Ibbotson Associates.

Ibbotson, Roger G., and Peng Chen. 2003. "Long-Run Stock Returns: Participating in the Real Economy." *Financial Analysts Journal* 59 (1): 88–98. doi:10.2469/faj.v59.n1.2505.

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 James P. Harrington. 2020. *Stocks, Bonds, Bills, and Inflation® (SBBI®): 2020 Summary Edition*. Charlottesville, VA: CFA Institute Research Foundation. https://rpc.cfainstitute.org/research/foundation/2020/sbbi-2020-summary-edition.

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

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.

Ilmanen, Antti. 2011. *Expected Returns: An Investor's Guide to Harvesting Market Rewards*. Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118467190.

Ilmanen, Antti, and Thomas Maloney. 2026. *Understanding Return Expectations*. Charlottesville, VA: CFA Institute Research Foundation.

Lintner, John. 1965a. "Security Prices, Risk, and Maximal Gains from Diversification." *Journal of Finance* 20 (4): 587–615. doi:10.2307/2977249.

Lintner, John. 1965b. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47 (1): 13–37. doi:10.2307/1924119.

McQuarrie, Edward F. 2024. "Stocks for the Long Run? Sometimes Yes, Sometimes No." *Financial Analysts Journal* 80 (1): 12–28. www.tandfonline.com/doi/full/10.1080/0015198X.2023.2268556.

Mehra, Rajnish, and Edward Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15 (2): 145–61. doi:10.1016/0304-3932(85)90061-3.

Mossin, Jan. 1966. "Equilibrium in a Capital Asset Market." *Econometrica* 34 (4): 768–83. doi:10.2307/1910098.

Perold, André F. 2004. "The Capital Asset Pricing Model." *Journal of Economic Perspectives* 18 (3): 3–24. doi:10.1257/0895330042162340.

Pfau, Wade. 2011. "Are TIPS Really Safe and Worry-Free?" *Advisor Perspectives* (15 November). www.advisorperspectives.com/articles/2011/11/15/are-tips-really-safe-and-worry-free.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* 19 (3): 425–42. doi:10.2307/2977928.

Shiller, Robert J. 2000. *Irrational Exuberance*. Princeton, NJ: Princeton University Press.

Siegel, Jeremy J. 1994. *Stocks for the Long Run*. New York: McGraw-Hill.

Siegel, Laurence B. 2017. *The Equity Risk Premium: A Contextual Literature Review*. Charlottesville, VA: CFA Institute Research Foundation.

———. 2018. "She Caught the CATY and Left Me a Mule to Ride: Improving on Dividend Yield as an Indicator of Stock Valuations and Expected Returns." AJO Vista (15 March). https://larrysiegel.org/she-caught-the-caty.

Straehl, Philip U., and Roger G. Ibbotson. 2017. "The Long-Run Drivers of Stock Returns: Total Payouts and the Real Economy." *Financial Analysts Journal* 73 (3): 32–52. doi:10.2469/faj.v73.n3.4.

Treynor, Jack L. 1962. "Toward a Theory of Market Value of Risky Assets." Unpublished manuscript. Final version in *Asset Pricing and Portfolio Performance*, edited by Robert A. Korajczyk (1999), 15–22. London: Risk Books.

Wilson, Jack W., and Charles P. Jones. 2002. "An Analysis of the S&P 500 Index and Cowles' Extensions: Price Indexes and Stock Returns, 1870–1999." *Journal of Business* 75 (3): 505–33. doi:10.1086/339903.