Dynamic Asset Allocation Using Machine Learning: Seeing the Forest for the Trees

Christian Mueller-Glissmann and Andrea Ferrario

Christian Mueller-Glissmann
is a managing director and head of Asset Allocation Research at Goldman Sachs in London, UK.
christian.mueller-glissmann@gs.com

Andrea Ferrario
is an associate in the Asset Allocation Research team at Goldman Sachs in London, UK.
andrea.ferrario@gs.com

KEY FINDINGS

- The authors construct three scores that track growth, inflation, and policy along the US business cycle. These scores combine multiple macro and market variables to improve their signal-to-noise ratio and have history starting in 1950.
- A machine learning algorithm (random forest) uses these scores combined with a risk appetite indicator to identify which macro regimes drive specific tail risks, such as large drawdowns or rallies for 60/40 portfolios. The model captures the outcomes driven by inflationary regimes well but struggles to capture growth inflection points.
- Several dynamic asset allocation overlays for a US 60/40 portfolio are then implemented based on the estimated probabilities of different market outcomes. The overlays materially enhanced risk-adjusted returns compared with a static 60/40 portfolio since 1950, although results were mixed over time.

ABSTRACT

High inflation and aggressive monetary policy tightening in 2022 triggered one of the largest return drawdowns for a US 60/40 portfolio in the last 100 years. In this article, the authors develop a dynamic asset allocation framework based on macro regimes using machine learning to improve the risk/reward versus static balanced portfolios with higher macro volatility. Using both macro and market data, they construct indicators for growth, inflation, and policy to track the business cycle and for risk appetite since 1950. They then use a random forest algorithm on those indicators to identify macro regimes that drive tail risks that matter for portfolio construction around a US 60/40 equity/bond portfolio. Based on real-time regime probabilities, they implement one of three dynamic asset allocation overlays: 1) switch between a 60/40 portfolio and cash, 2) rotate between equities and bonds, and 3) allocate to commodities/gold. The overlays materially enhanced risk-adjusted returns compared with a static 60/40 portfolio since 1950, although results were mixed over time.

A buy-and-hold 60/40 equity/bond portfolio in the United States performed strongly for the last generation of investors, but the strategy was less successful in 2022 as high inflation and aggressive monetary policy tightening triggered one of the largest return drawdowns in the last 100 years. Higher macro volatility, especially inflation volatility, increases the value of dynamic asset allocation approaches relative to static 60/40 portfolios. This article adds to the stream of literature aiming to identify macro regimes to dynamically adjust portfolios and improve their risk/reward through the cycle (see, for example, Ang and Bekaert 2004; Sheikh and Sun 2012; Nystrup et al. 2015; Nystrup et al. 2017; and Uysal and Mulvey 2021).

To track the US business cycle, we construct three scores for growth, inflation, and policy since 1950, combining multiple macro and market variables to improve their signal-to-noise ratio. We also construct a risk appetite indicator to track extreme swings in investor sentiment along the business cycle. Cross-asset performance is closely linked to business cycle swings. Levels and changes in our indicators provide insight into the asymmetry for subsequent cross-asset returns. Looking at the business cycle along multiple dimensions allows us to better capture the relationship between macro and markets. Indeed, this relationship varies depending on the position in the cycle and levels of risk appetite as well as on the interactions among growth, inflation, and policy.

We use a machine learning algorithm, a so-called random forest, to identify macro regimes driving specific market outcomes based on the business-cycle scores and risk appetite indicator. Compared with linear approaches, this approach can better capture interactions and nonlinearities between the indicators and markets. It also allows us to define the market outcomes we want to investigate and track their probability in real time. We focus on important tail outcomes for balanced portfolios: left-tail and right-tail risk for 60/40 portfolios, equities markedly outperforming or underperforming bonds, and commodities/gold significantly outperforming 60/40 portfolios. We employ a variety of methods to visualize the inner workings of the random forest and outputs using Shapley values and Student trees. We find that the impact of the business-cycle scores on cross-asset performance is not symmetric: for example, high inflation levels and non-recessionary growth matter most for 60/40 drawdown risk, while sharp 60/40 recoveries tend to come from low levels of risk appetite among investors.

We use the probabilities estimated by our machine learning model based on the cycle scores to implement several dynamic asset allocation (DAA) overlays for a 60/40 portfolio. The overlays are as follows:

1. switch between a 60/40 portfolio and cash,
2. rotate between equities and bonds, and
3. allocate to commodities/gold.

We show that these DAA strategies would have enhanced risk-adjusted returns relative to a buy-and-hold 60/40 portfolio, even without the use of leverage or short positions. They work well for inflationary regimes but struggle to capture structural changes and growth inflection points when markets often lead macro.

TRACKING THE BUSINESS CYCLE AND ASSET PERFORMANCE

The performance of a US 60/40 portfolio varied materially with the business cycle. Growth, inflation, and policy (fiscal and monetary) are linked but often diverge, and equities and bonds tend to react differently to each of them. Normally, early cycle, as the economy recovers, growth shifts above trend without inflationary pressures while, late cycle, inflation tends to pick up. This eventually drives policy tightening, which pushes up bond yields and drives deleveraging as growth slows and shifts below trend, often ending in a recession. Bonds tend to outperform equities early in a recession before a usually sharp equity reversal as markets anticipate the recovery. Equities deliver the best risk-adjusted returns with low and anchored inflation and strong growth; these so-called Goldilocks regimes are also best for 60/40 portfolios.

Structural factors can result in very different business cycles and cross-asset performance. For example, owing to disinflationary pressures from technology and shale oil last cycle, growth often picked up with anchored inflation, which resulted in more time spent in Goldilocks regimes in which both equities and bonds can do well. On the flipside, during the 1970s "Stagflation" period, inflation was often elevated, with poor growth on account of oil supply shocks. Such inflationary regimes also affected the relationship between equities and bonds: with low and anchored inflation, equity–bond correlations tend to be negative as growth becomes a more important driver of bonds and monetary policy.

To decompose and compare the US business cycles, we construct growth, inflation, and policy scores using both macro and market data, as shown in Exhibit 1. These three scores help capture cycle differences and regimes, which matter for markets. For example, since the 1990s late-cycle regimes generated less inflation, which supported 60/40 portfolios. In addition, broad policy had larger disconnects with inflation, that is, central banks tightened policy without inflationary pressures, whereas during the COVID-19 recovery, central banks reacted slowly to the acceleration in inflation.

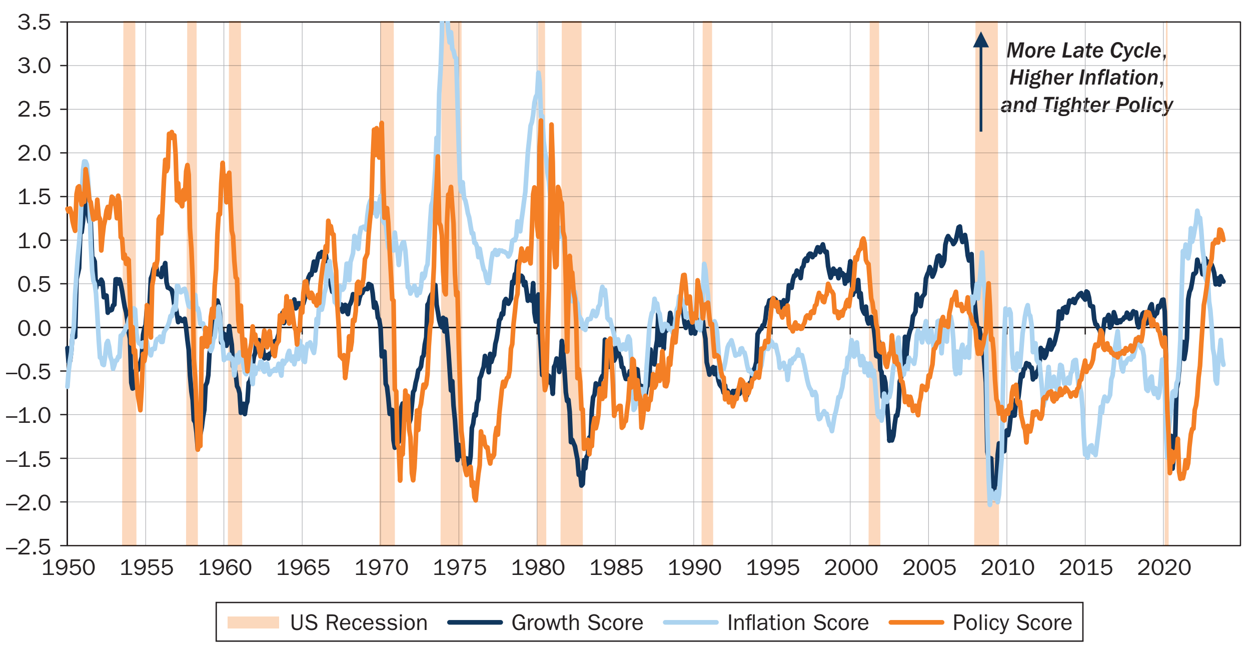
We only include US data given our focus on US 60/40 portfolios and historical data availability, and we combine different macro and market variables in an attempt to improve the signal-to-noise ratio and mitigate the lag in the availability of business-cycle data. We use long-term, expanding *z*-scores to track the cycle position and one-year rolling *z*-scores to understand macro momentum; a combination of both provides better signals for asset performance—for example, when growth momentum turns negative in a late-cycle backdrop (see the appendix for details).

Growth Cycles and Asset Performance

To track the growth cycle, we aggregate such macro variables as the US unemployment rate, output gap, private sector financial balance, and corporate margins, as well as cyclical risk premia in markets (see the appendix for details). Cyclical risk

EXHIBIT 1

Average Expanding z-Score of US Macro and Market Variables across Growth, Inflation, and Policy



SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

premia, for example equity and credit risk premia, tend to increase in recessions and decline into a late-cycle backdrop owing to supportive macro conditions. While we do not include global growth variables, they will in part be captured by US cyclical risk premia. The growth score level (expanding *z*-score) indicates whether the US economy is early, mid, or late cycle or in a recession. The higher the level, the later in the cycle the economy is. The growth score tends to trough around the end of a recession. Growth momentum (one-year *z*-score) inflects earlier, potentially helping to capture turning points (see the appendix).

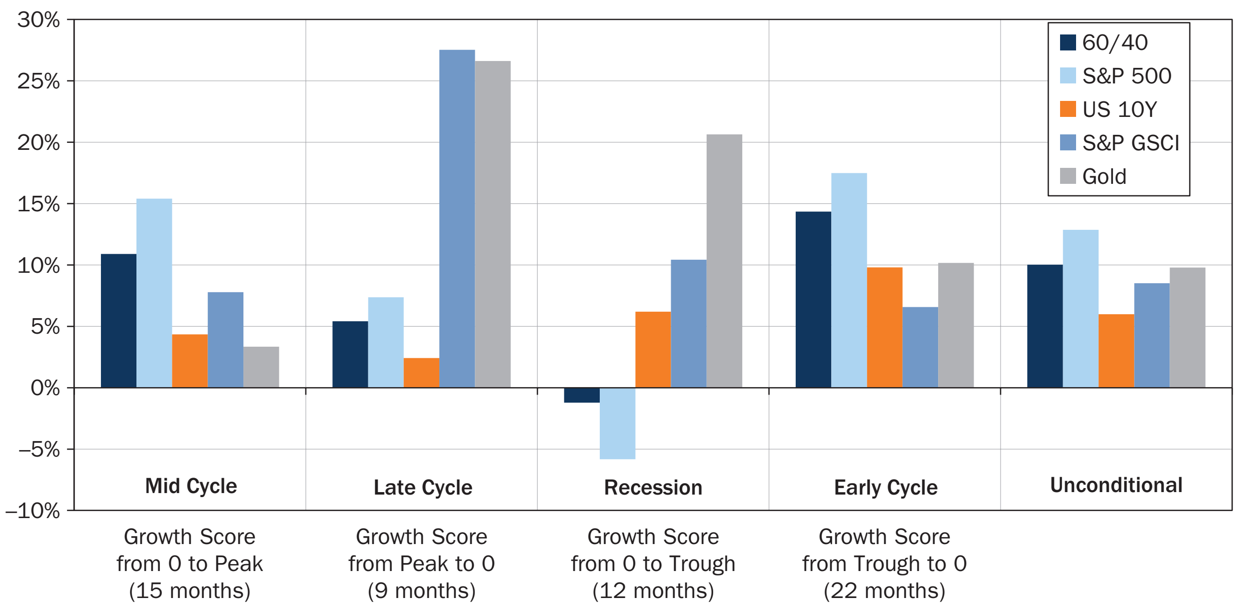
As Exhibit 2 shows, 60/40 portfolios suffer most when growth score levels are negative and falling, which tends to be during recessions. Such 60/40 sell-offs were driven by equities, whereas bonds, commodities and gold on average outperformed a 60/40 portfolio in that regime. Instead, from a trough in the growth score level equities performed best on average. Commodities performed best mid- and late cycle, when growth was strong. Bond returns were best during recessions and early-cycle episodes, when inflation was anchored and policy was easing.

At extremes, the growth score level conditioned subsequent equity and bond returns. During late-cycle backdrops (growth score above 0.75), right-tail risk for equities is lower and returns are more negatively skewed, while early cycle (a growth score below −0.75) the reverse is true (Exhibit 3, Panel A). In particular, risk premia are often high around low growth score levels, creating potential for a strong recovery in equities. For bonds, the link to growth is less symmetric, with yields generally skewed to the downside when late cycle but less affected otherwise (Exhibit 3, Panel B).

As Panel A in Exhibit 4 shows, while equities generally consolidated post a peak in the growth score, there was a large performance dispersion. We found that the interaction of growth with inflation and policy mattered a lot. And there were wrong signals: During the mid-1960s and mid-1990s, the growth score was elevated but

EXHIBIT 2

Average 12-Month Total Returns for Different Growth Regimes

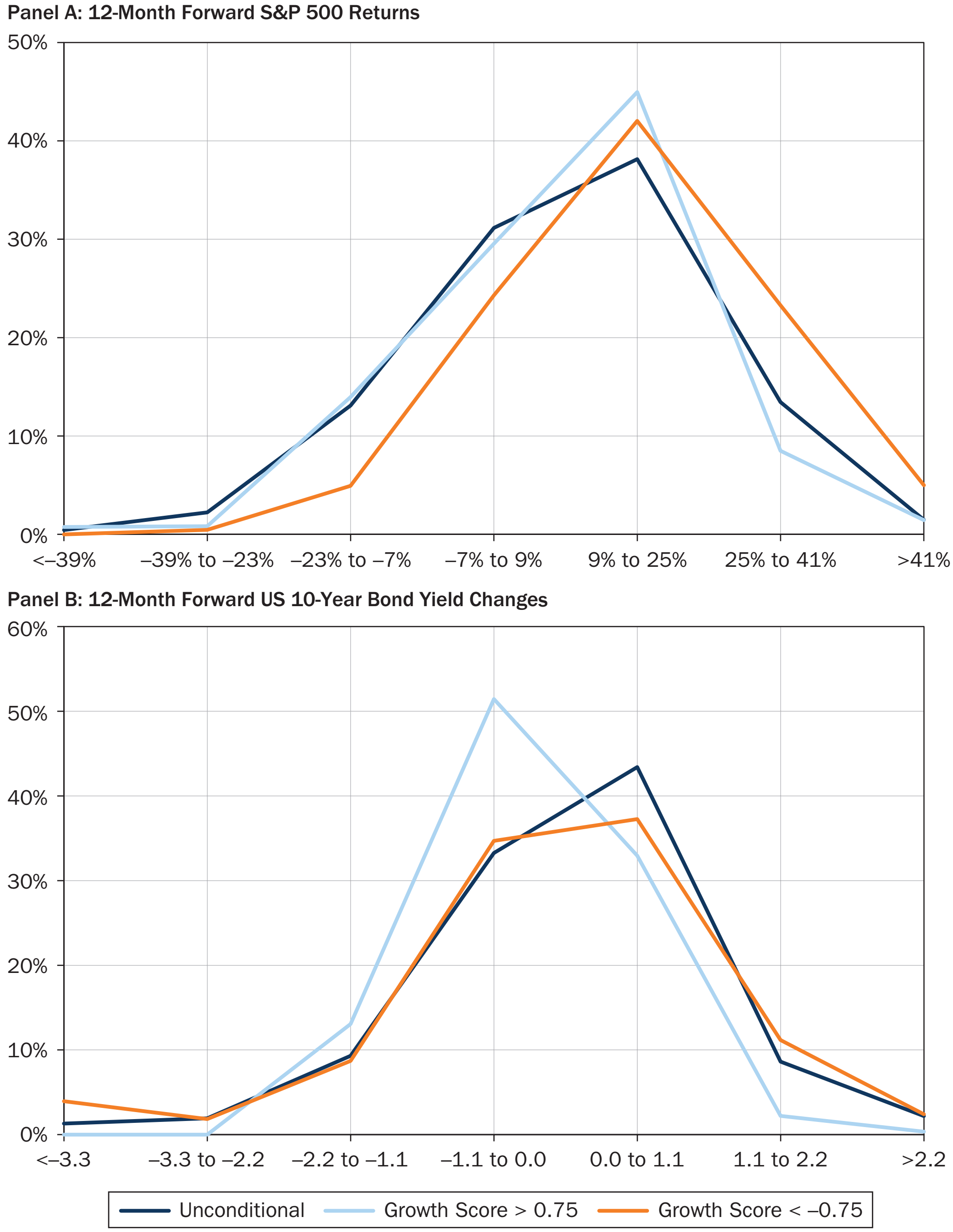


NOTE: Median period length is shown in brackets; data for January 1950–November 2023.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

EXHIBIT 3

Distribution of 12-Month Forward S&P 500 Index Returns and US 10-Year Bond Yield Changes from Different Growth Score Levels, January 1950–November 2023

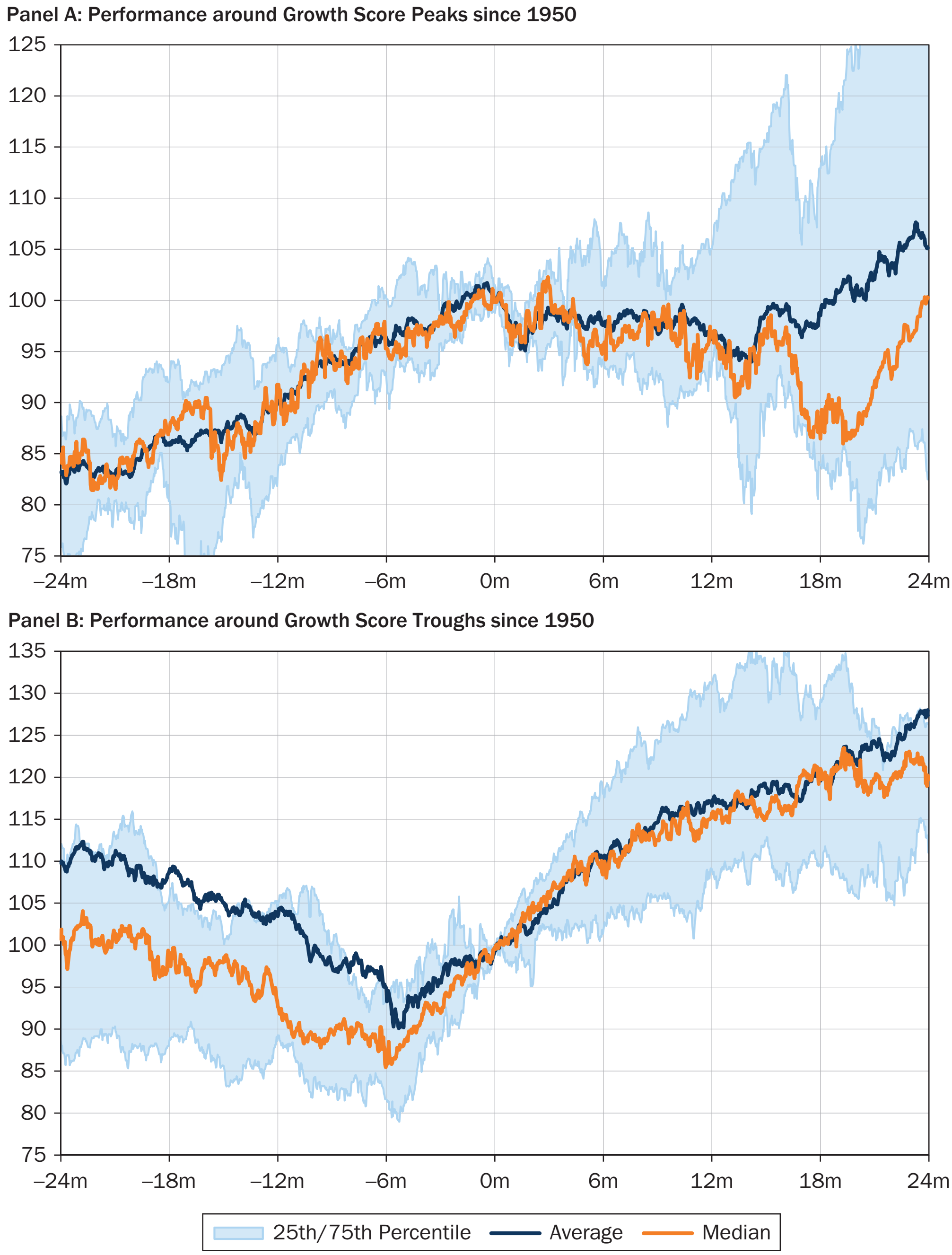


SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

declined without a recession and equities performed well, boosted by structural tailwinds. And, even if there was a subsequent recession, equities often performed well before it took hold. As such, timing equities around recessions is particularly hard. Furthermore, as Panel B in Exhibit 4 shows, markets tend to lead growth data both during recessions and in recoveries. Equities recovered, on average, six months before the growth score levels troughed.

EXHIBIT 4

Average S&P 500 Performance around Growth Score Peaks and Troughs since 1950 (rebased to 100, data January 1950–November 2023)



SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

The Impact of Inflation and Policy

To track inflation along the US business cycle, we combine reported measures, such as CPI, PPI, and PCE, with market variables, including commodity prices and breakeven inflation (see the appendix for details). Again, we look at long-term expanding *z*-scores to assess the inflation cycle position and one-year rolling *z*-scores for momentum. The 1970s and the recovery from the COVID-19 crisis were the main high-inflation periods, but there have been several smaller inflation cycles. Market and macro variables for inflation have been closely linked but market pricing often leads. In particular, the price of oil, which is a key driver of inflation, tends to have large swings. For this reason, commodities and, in particular, oil have been successful hedges for inflation.

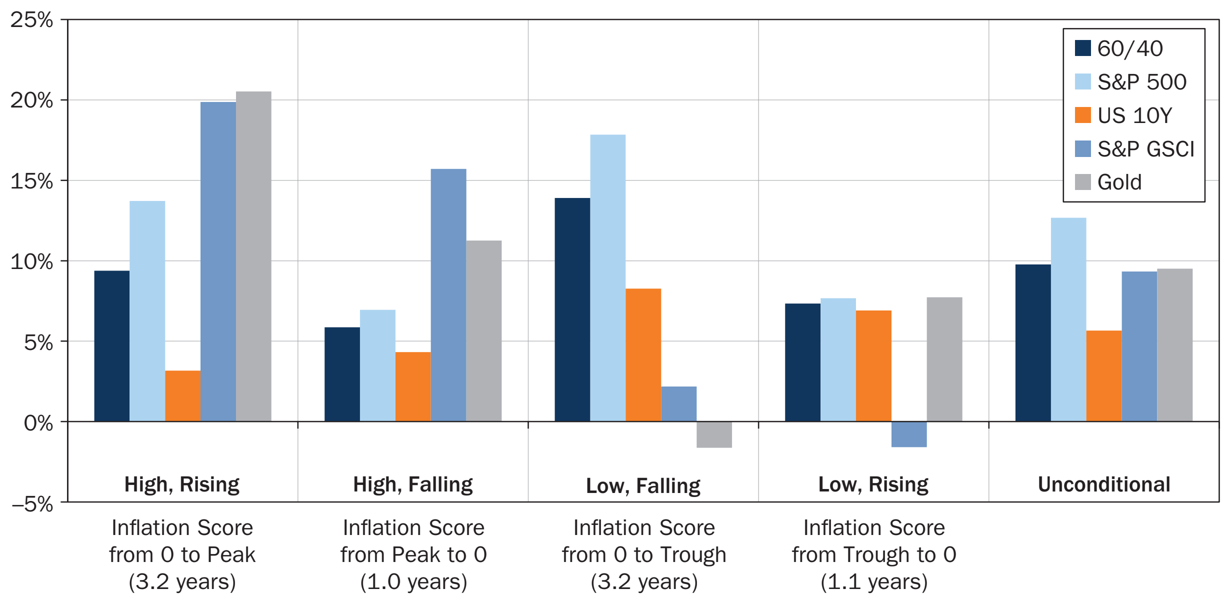
Inflation cycles generally lasted longer than growth cycles and were more persistent. As Exhibit 5 shows, a high inflation score has boosted commodities and gold, especially with rising inflation. Bonds suffer with positive inflation levels and momentum. Equities have performed best when inflation scores have been low and falling, which is also when 60/40 portfolios performed best. For equities, the interaction with growth and policy is critical; for example, in 2022, equities suffered alongside bonds owing to elevated inflation and policy tightening, whereas in 2021 they performed well despite rising inflation in the second half of the year.

While equities have a less clear link to inflation, elevated inflation can limit right-tail risk and increase left-tail risk. This likely owes to impending policy tightening and often inflation tends to be higher in a late-cycle position. On the flipside, low inflation scores increase right-tail risk for equities, in part on account of recovery potential after recessions. High inflation increases right-tail risk for US 10-year bond yields, but low inflation does not materially change the forward distribution of yield changes.

To track policy, we combine macro variables such as the US Federal Reserve's funds rate versus its 10-year average, M2 money supply, and the US fiscal deficit, as well as such market measures as the yield curve and shorter-dated real yields (see the appendix for details). Policy can shift quickly and through multiple channels, such as during the COVID-19 crisis, when both fiscal and monetary policy eased sharply to prevent a deeper global recession. While inflation and policy should be closely linked because central banks target inflation, during the COVID-19 recovery, there was a large lag in policy. More recently, policy tightening has continued despite a disinflationary trend. Moreover, policy has been more "shocking" compared with inflation, which built

EXHIBIT 5

Average 12-Month Total Returns for Different Inflation Regimes



NOTE: Median period length is shown in brackets; data from January 1950–November 2023.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

up gradually and was persistent, at least in the 1970s. During recessions, policy has often shifted aggressively, buffering the growth cycle.

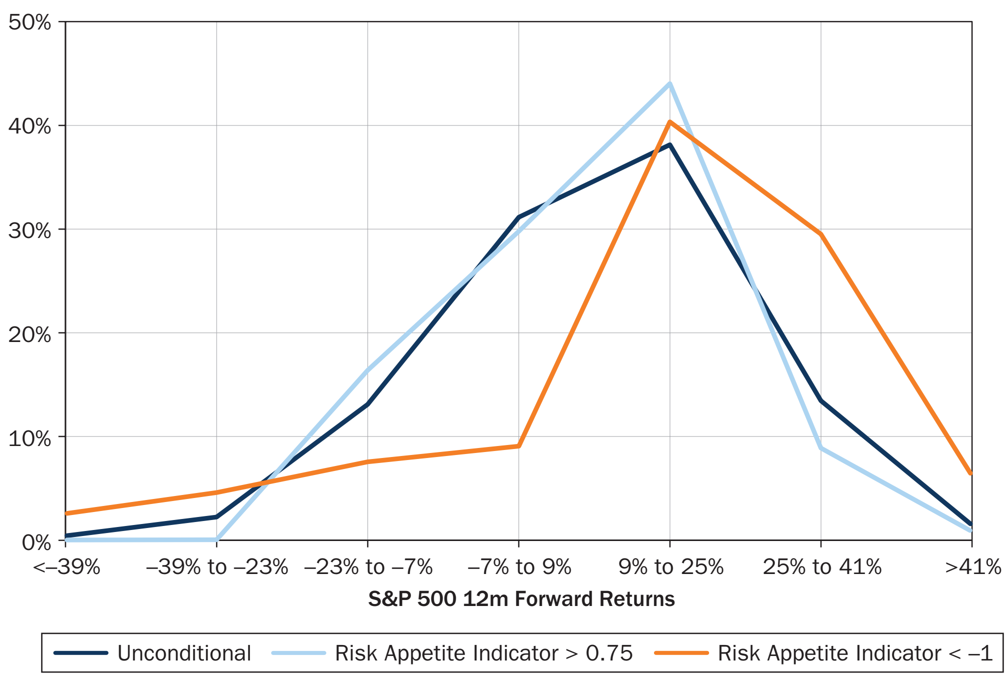
We found the impact of either extremes in policy alone—tightening or easing—has had relatively smaller impacts on the 12-month forward return distributions for equities. In periods of very tight policy, 12-month forward returns were slightly negatively skewed, whereas very easy policy drove more right-tail risk. In periods of policy tightening, however, US 10-year bond yields in the next 12 months were somewhat positively skewed and the reverse was true for policy-easing periods. While the direct impact was more limited, the interaction of policy with both growth and inflation is critical, especially during recessions.

Incorporating Extremes in Investor Risk Appetite

Markets tend to be forward looking, and as a result, equities lead the business cycle, in particular growth data. To better capture reversals, we construct a risk appetite indicator (RAI), which captures how bullish or bearish investors are. We use the average of one-year rolling *z*-scores of the total return performance of the S&P 500 Index versus US 10-year bonds, the relative performance of US cyclical versus defensive equities, as well as USD BAA credit spreads to proxy broad risk appetite (see the appendix for details). The RAI is supposed to help with reversals at extremes in investor sentiment; for example, during recessions when investors are already very bearish, downside risks for equities start to be limited and they often recover before a recession ends. Or if the macro backdrop is supportive for risky assets but investors are already bullish, there is greater risk of disappointment. Indeed, as Exhibit 6 shows, at RAI levels below –1, the 12-month forward return distribution for

EXHIBIT 6

Distribution of 12-Month Forward S&P 500 Returns from Different RAI Levels (data from January 1950–November 2023)



SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

the S&P 500 was more positively skewed with more right-tail risk. On the flipside, when the RAI was above 0.75 (i.e., investors were bullish), upside risk for equities was increasingly capped even though left-tail risk was not much more elevated compared with the unconditional distribution.

SEEING THE FOREST FOR THE TREES: IDENTIFYING MACRO REGIMES

As documented by Bartram, Branke, and Motahari (2020) and Lee et al. (2023), interest in and applications of machine learning for asset management have increased over the past years. The literature on AI/machine learning for portfolio management includes studies focused on predicting expected returns (e.g., for stock selection and factor investing), sentiment analysis, predicting variances/covariances, portfolio optimization, as well as hedging and trading strategies. We use a machine learning algorithm, a so-called random forest, to identify macro regimes and subsequent market outcomes. In short, this algorithm estimates multiple decision trees to see which combinations of cycle scores and risk appetite indicator drive specific cross-asset returns over time. Each tree is estimated on a random subset of the training sample and searches for the best threshold in each node among a random subset of the features. Because this procedure results in a greater diversity among the trees, their prediction errors tend to cancel out, generally yielding an overall better performance of the aggregated result (see Ho 1995; Breiman 2001; and Géron 2022 for a broad reference on random forest models).

One of the advantages of random forests compared with other machine learning and regime classification models is that it is a "supervised" algorithm. This means that we can define the market outcomes we want to forecast, which allows us to focus on specific outcomes that are most important for balanced portfolio construction. By contrast, "unsupervised" algorithms, such as Gaussian mixture, Wasserstein k-means clustering, and hidden Markov models, attempt to find clusters of similar observations among macroeconomic or market variables. Moreover, random forests can capture nonlinearities between market and macro data and interaction effects across the cycle scores better than simpler linear models. For example, a late-cycle position without inflation, policy tightening, or negative growth momentum is less concerning than one with overheating and slowing growth. In a similar prediction exercise, Yazdani (2020) found that a random forest achieves the best performance in forecasting recessions among competing linear and machine learning models. That said, Blitz et al. (2023) highlight that applications of machine learning to financial markets are susceptible to some pitfalls, such as low signal-to-noise ratios and smaller datasets. We try to mitigate those by aggregating multiple macro and market series into our cycle scores to reduce noise (see the previous section for more details) and use a long history dataset extending back to 1950 to encompass multiple business and structural cycles.

As inputs, we use levels (back-expanding *z*-scores) and momentum (one-year rolling *z*-scores) of our growth, inflation, and policy scores, as well as the risk appetite indicator to capture extremes in investor sentiment. The dependent variable is a binary dummy representing a corner outcome for a 60/40 portfolio or cross-asset performance in the following 12 months (e.g., a 60/40 12-month forward return below 0%).[1]

[1] Because we focus on the tails of the return distribution, the outcomes of the dependent variable are unbalanced (e.g., 60/40 12-month returns are negative only 17% of the time); this can lead to a biased random forest estimator. To adjust for this, each outcome is associated with a weight inversely proportional to its frequency when fitting the model. Lastly, the fitted probability from the random forest is adjusted to reflect the true unconditional probability of the outcome using the procedure in Elkan (2001).

To avoid overfitting the in-sample data, we constrain the maximum depth of each decision tree to two layers and set a minimum size for the final leaves.[2]

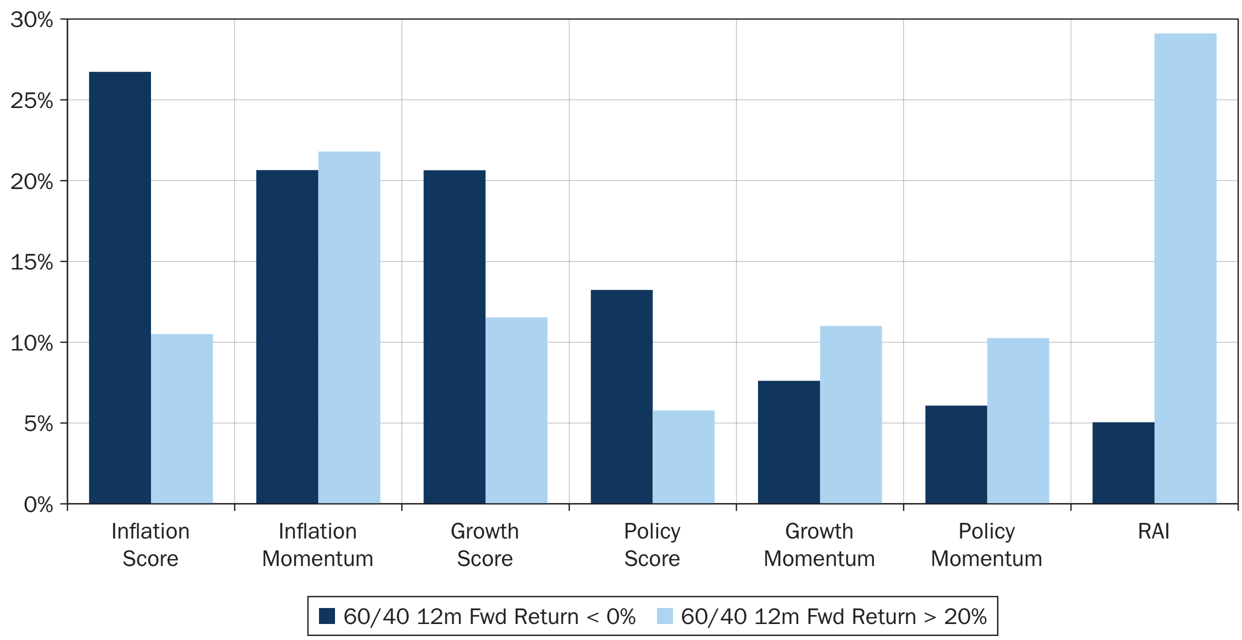
Likelihood of 60/40 Drawdowns/Recoveries

First, we assess tail risk for 60/40 portfolios over the next 12 months. Avoiding large 60/40 drawdowns can materially enhance returns, but we also try to capture subsequent recoveries. Exhibit 7 highlights the cycle indicators that matter most for regimes with elevated 60/40 tail risk (12-month returns below 0% or above 20%). For large drawdowns, inflation, growth, and policy scores, as well as inflation momentum mattered most; for strong returns, the level of risk appetite and inflation momentum were most important.

To illustrate the drivers of regimes, we can create a so-called Student tree that summarizes the main relationships captured by the random forest (see also Zhou et al. 2018). For example, as shown in Exhibit 8, Panel A, the likelihood of a large 60/40 drawdown increases late cycle with an elevated inflation score coupled with a non-recessionary growth score, as that combination might result in a higher chance of overheating. If the inflation score is negative, 60/40 portfolios can do well

EXHIBIT 7

Importance of Cycle Scores in Determining Forward 60/40 Returns



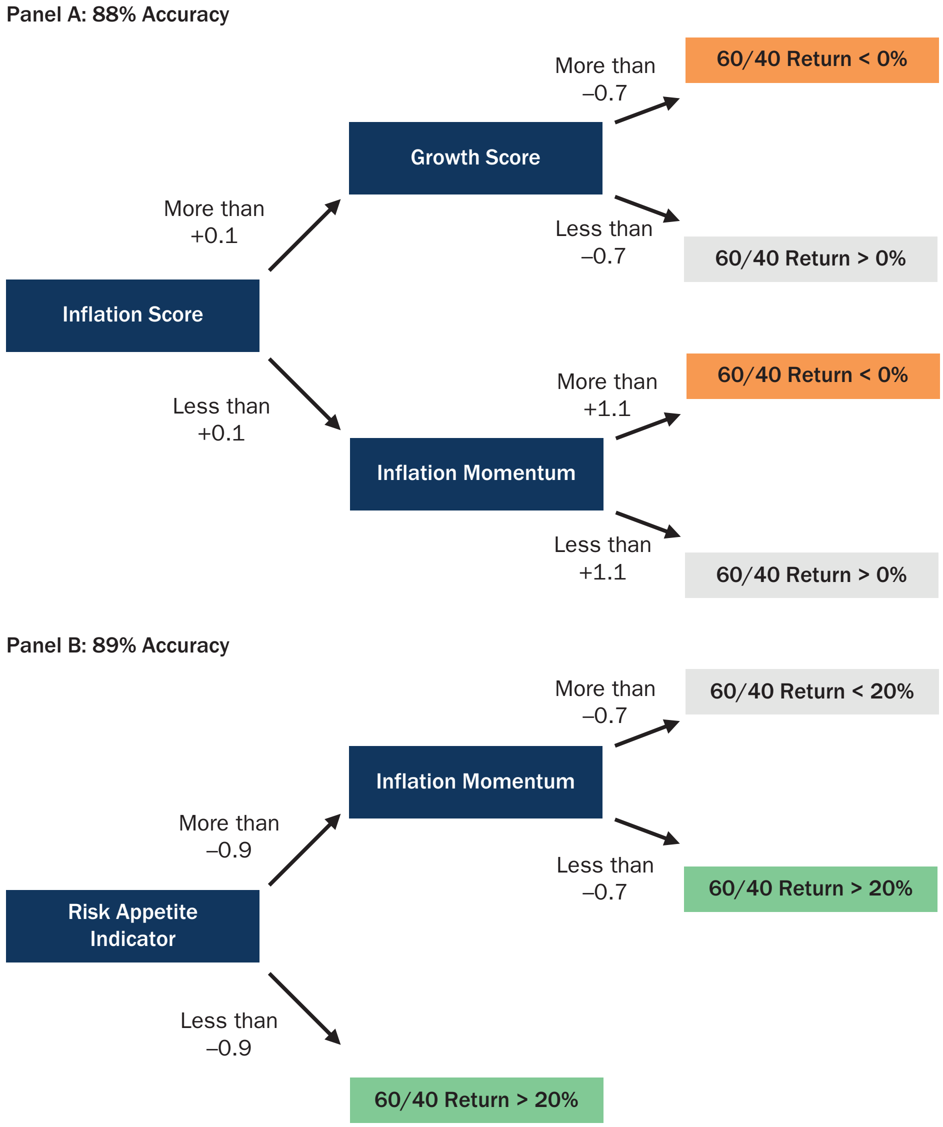
NOTES: Relative importance adds to 100%. Model fitted on a January 1950–December 2021 sample.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

[2] The random forest models are estimated in sample with data from 1950 to 2021 for business-cycle scores and to 2022 for corresponding asset performance. We also assess the out-of-bag performance of the random forest models over the same period and find results consistent with the in-sample performance. As a further check, we estimate the random forest models from 1950 to 2018 and check their out-of-sample performance since 2020. While the out-of-sample interval is relatively small, we find that the classification performance metrics are in line with, if not better than, the in-sample one for most models.

EXHIBIT 8

Student Tree with 88% and 89% Accuracy in Approximating the Original Random Forest Model



NOTE: Fit on data for January 1950–December 2021.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

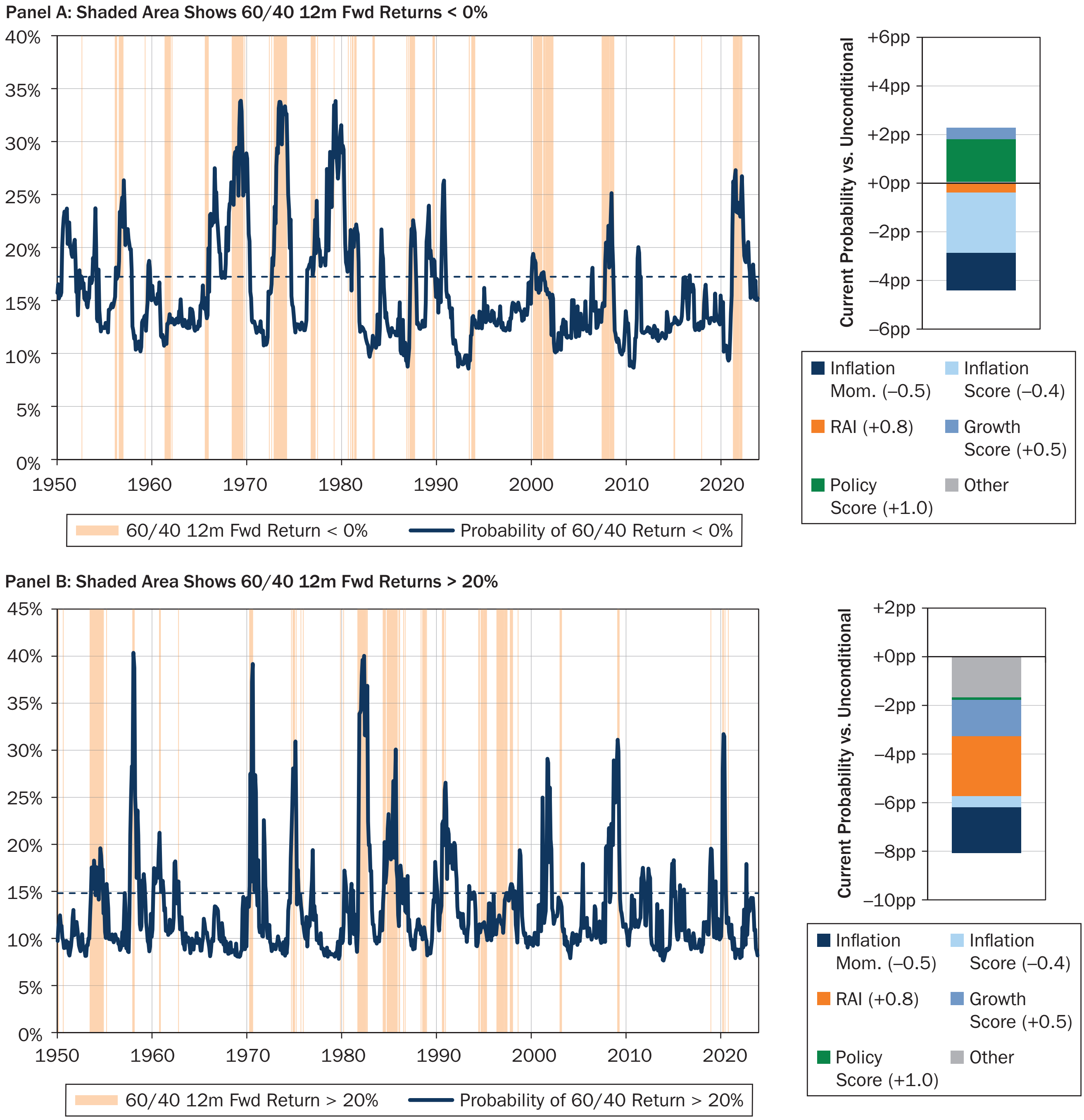
independent of growth (in part because bonds can buffer equities), but if it is positive, the risk of rising bond yields weighing on equities is high.

On the flipside, as shown in Panel B of Exhibit 8, the potential for a large rally for a 60/40 portfolio is high if the risk appetite indicator is very low (likely as there has just been an equity bear market/recession) or if inflation momentum is negative, likely with both equities and bonds boosted by falling bond yields. If the risk appetite indicator is positive, unless inflation momentum is very negative, right-tail risk for 60/40 portfolios is capped.

Based on the random forest we can estimate the probability of the 60/40 tail risk regimes in real time. Exhibit 9 shows the model captured several of the large 60/40 drawdowns and rallies. Inflationary 60/40 drawdowns were anticipated better, while the probability did not pick up materially ahead of the collapse of the Tech Bubble

EXHIBIT 9

Random Forest Probability Estimates of US 60/40 Portfolio, January 1950–November 2023



NOTES: Dotted line denotes unconditional probability. The bar chart shows the Shapley decomposition of the current probability (latest score in brackets).

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

and the Global Financial Crisis (GFC). This is because of, at least in part, specific structural imbalances (equity and credit bubbles respectively) that were not captured as well by our cycle scores.

The 60/40 recovery probability also increased too early during the collapse of the Tech Bubble and the GFC. This was in large part on account of the depth of those bear markets, which meant that the drawdown continued even after the risk appetite indicator was already very depressed. Recently, the model has indicated falling 60/40 drawdown risk, but the risk of a large 60/40 rally is now also more limited after a strong recovery in 2023. On the right-hand side of the panels in Exhibit 9, we show a Shapley decomposition with the drivers of the probability versus unconditional;[3] recently, low inflation levels and momentum have reduced the risk of further 60/40 drawdowns in the next 12 months, but continued policy tightening increased the probability.

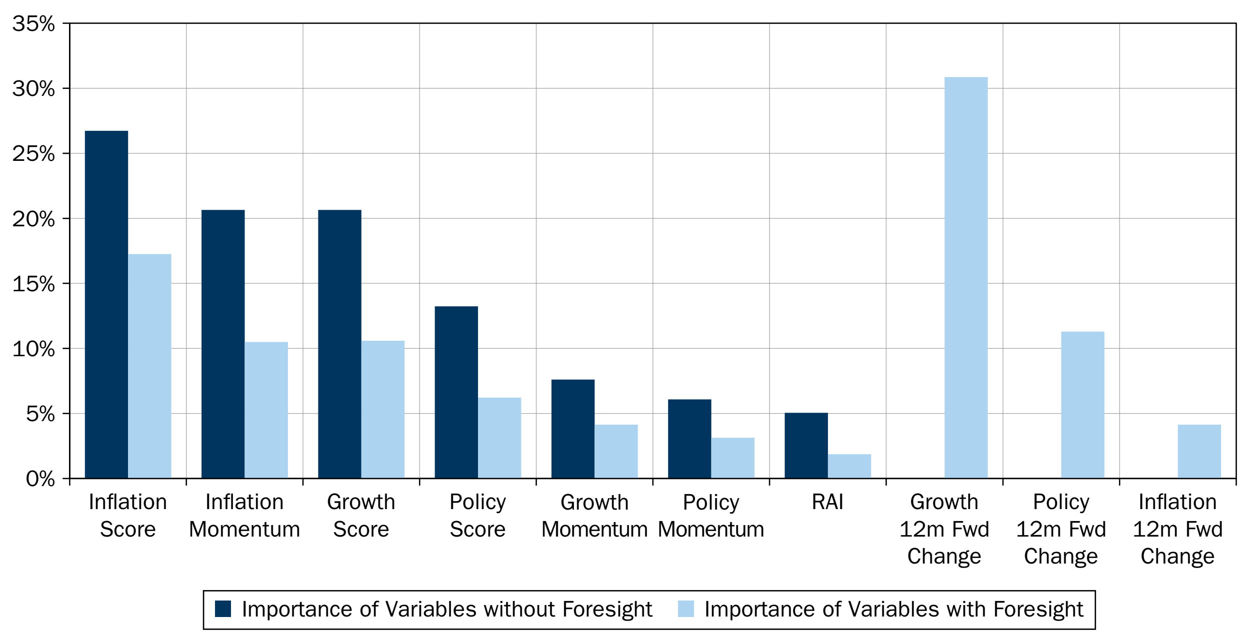
The current probability estimate is based on data available in real time. This often means lags for macro data releases relative to the period covered, and as a result, our probability estimates can be slower than markets at turning points. As macro regimes tend to cluster and there is mean reversion at extremes, it should be possible to capture regimes at least partially without the need for forecasts.

However, taking forward macro views can materially improve probability estimates; including the growth, inflation, and policy scores 12-months ahead with perfect foresight changes which variables are most important. Exhibit 10 shows that knowing growth in the next 12 months is the most useful, but inflation levels remain important for the risk of a 60/40 portfolio drawdown.

Exhibit 11 shows a comparison of the probability estimates with and without perfect foresight. With perfect foresight, the base-case probability is lower when there is no risk of a 60/40 drawdown, and it often increases more into and during actual drawdowns. Especially during the Tech Bubble and GFC drawdowns, the probability increased more and remained elevated for longer, providing a much better signal.

EXHIBIT 10

Importance of Cycle Scores in Determining Forward 60/40 Returns < 0% (including perfect foresight)



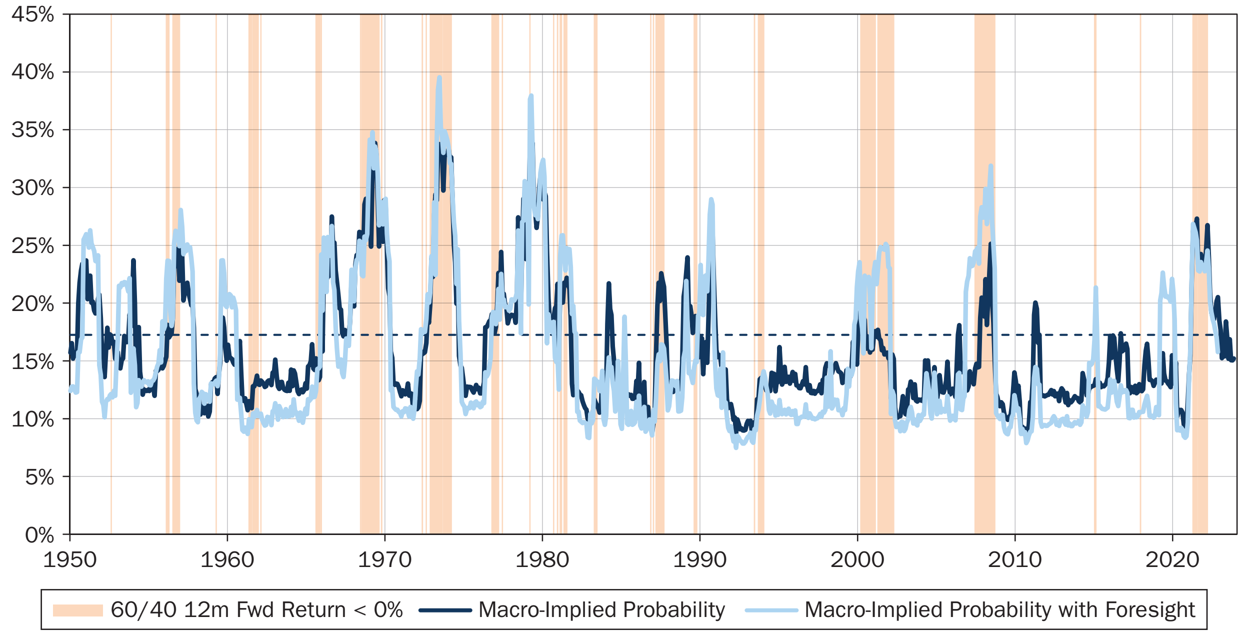
NOTE: Relative importance adding to 100%; model is fitted on a January 1950–December 2021 sample.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

[3] The contribution to the fitted probability is computed using the SHapley Additive exPlanations (SHAP) approach from Lundberg and Lee (2017), which is based on Shapley values.

EXHIBIT 11

Random Forest Probability Estimate of 60/40 Portfolio Drawdown with and without Perfect Foresight



NOTES: Dotted line denotes unconditional probability. Data for January 1950–November 2023.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

Likelihood of Equities Underperforming/Outperforming Bonds

Next, we assess the potential for equities to outperform bonds or vice versa to manage tail risk in 60/40 portfolios while staying invested. As Exhibit 12 shows, the most important drivers for the regime when equities underperform bonds by more than 10% in the next 12 months are similar to those for 60/40 drawdowns: growth, inflation, policy levels, and inflation momentum. However, the potential for equities to outperform bonds by more than 20% is closely linked to growth (level and momentum) as well as inflation levels.

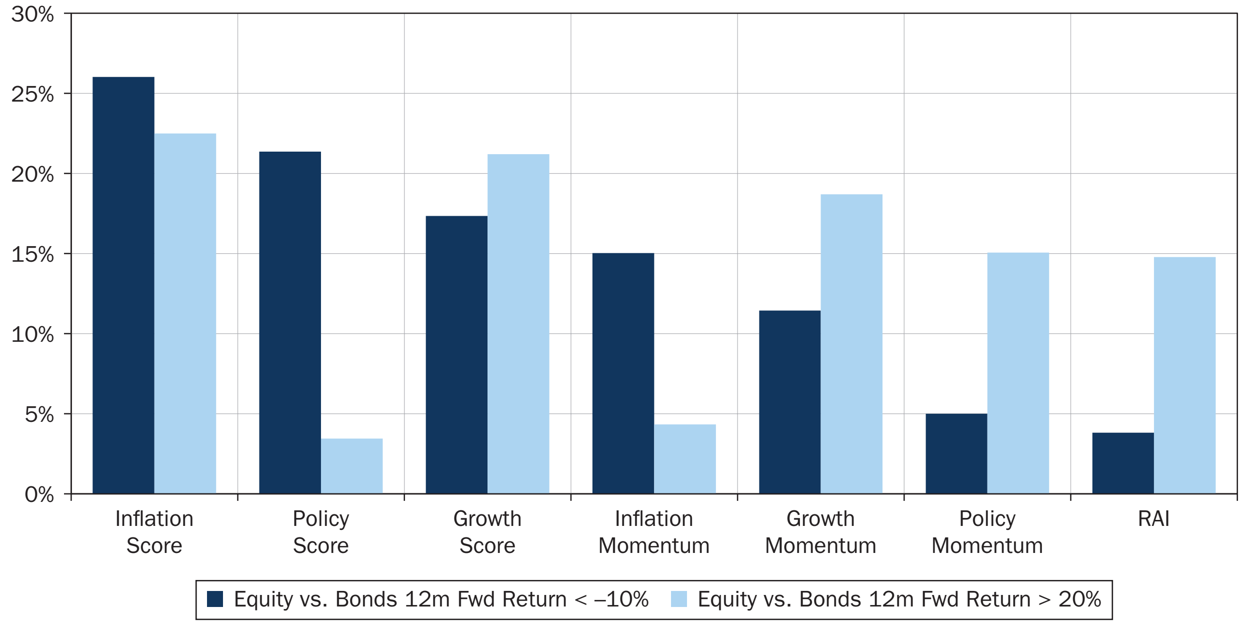
Very similar to the likelihood of a large 60/40 drawdown, a late-cycle backdrop coupled with elevated inflation tends to increase the risk of a large equity drawdown versus bonds, most likely owing to the potential for policy tightening and elevated recession risk. The likelihood of equities outperforming bonds increases with a lower growth level and momentum at the onset, which signal the potential for a market reversal; positive growth scores coupled with negative inflation scores (Goldilocks scenario) or very low risk appetite also increase that potential.

Exhibit 13 shows that the random forest probability estimate captured several of the regimes with large equity versus bond performance swings. The current probability of equities performing poorly versus bonds has declined in 2023. However, the likelihood of equities outperforming bonds materially also remains relatively low, in large part owing to the late-cycle backdrop.

However, once again there were wrong signals around market reversals during the Tech Bubble and GFC. Moreover, the model will not capture structural changes in the sample period, which can matter a great deal for equity versus bond performance. For example, equities might have become more or less sensitive to growth or rates owing to changing sector/style weights. Alternatively, if there is little inflation volatility, as has been the case since the 1990s, growth becomes a much more important driver.

EXHIBIT 12

Importance of Cycle Scores in Determining Forward Equity vs. Bond Performance Tails



NOTES: Relative importance adds to 100%. The model is fitted on a January 1950–December 2021 sample.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

Moreover, despite being late cycle, equities can still be boosted by long-term growth optimism, such as during the Tech Bubble.

Which macro data matter also varies over time with structural regimes. For example, using different training samples for the random forests results in very different drivers that were most important for the likelihood of equities underperforming bonds. As Exhibit 14 shows, before the 1990s inflation levels and momentum mattered most, especially in the 1970s when inflation volatility triggered rates and growth volatility, similar to 2022. However, from 1990 until 2021, when inflation was low and anchored, growth momentum was the key driver to equity drawdown risk, and growth volatility actually triggered rate volatility (resulting in negative equity/bond correlations).

Exhibit 15 shows that, as a result, random forest probability estimates were also very different: An estimate based on the 1950–1989 sample captured the inflationary equity bear markets well, whereas an estimate based on the 1990–2021 sample did a better job in anticipating the collapse of the Tech Bubble and the GFC. However, both did poorly for their respective opposite sample drawdowns and worked less well for the whole sample.

Likelihood of Commodities/Gold Outperforming a US 60/40 Portfolio

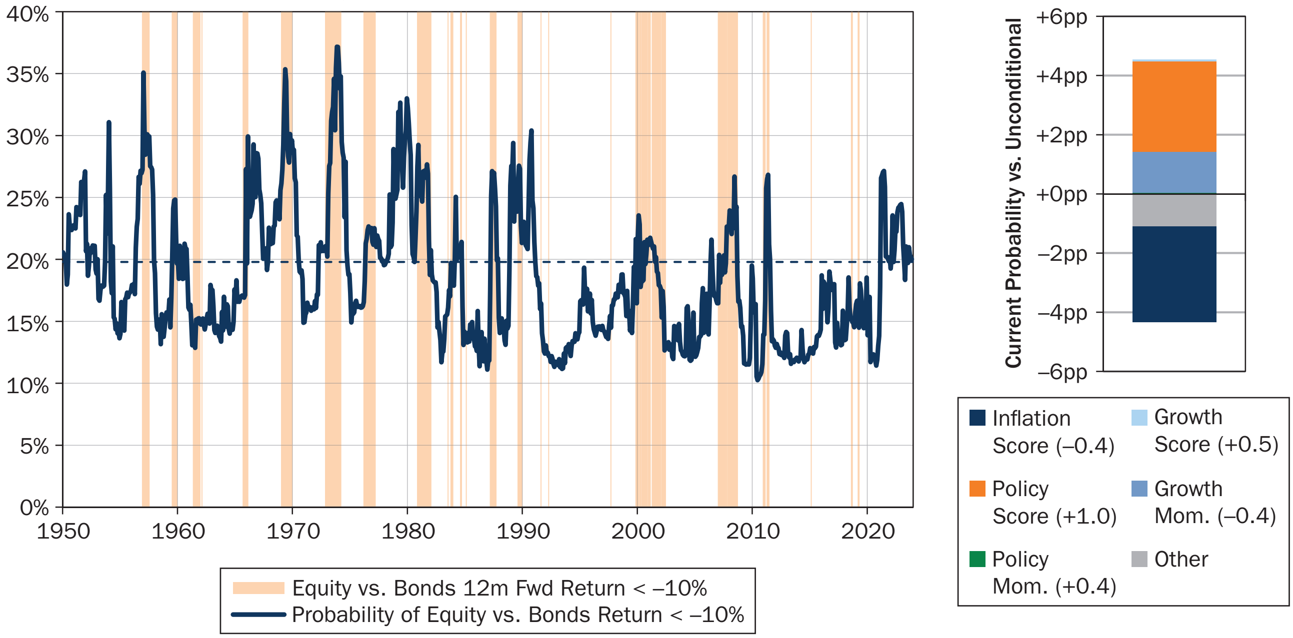
The third regime we try to capture is when commodities/gold materially outperform versus a 60/40 portfolio in the subsequent 12 months. This should help investors to dynamically allocate to those assets to diversify in periods of high inflation and protect real returns. As shown in Exhibit 16, for S&P GSCI outperformance most macro variables outside of policy were equally important, suggesting the breadth of signals mattered for the regime probability. For gold outperformance, the most important driver is the inflation level (higher is better).

Exhibit 17 shows that the model captured high inflation regimes, such as the 1970s and COVID-19 recovery, well, which triggered significant commodity outperformance

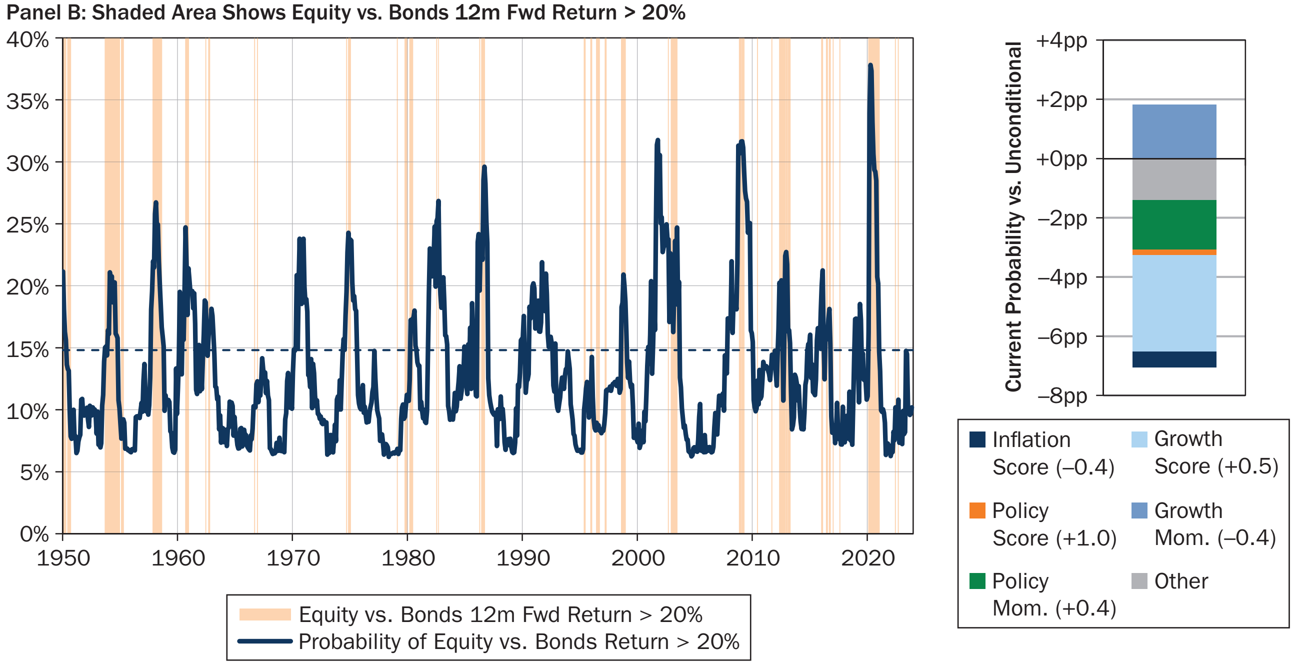
EXHIBIT 13

Random Forest Probability Estimate of Equity vs. Bonds Return

Panel A: Shaded Area Shows Equity vs. Bonds 12m Fwd Return < –10%



Panel B: Shaded Area Shows Equity vs. Bonds 12m Fwd Return > 20%



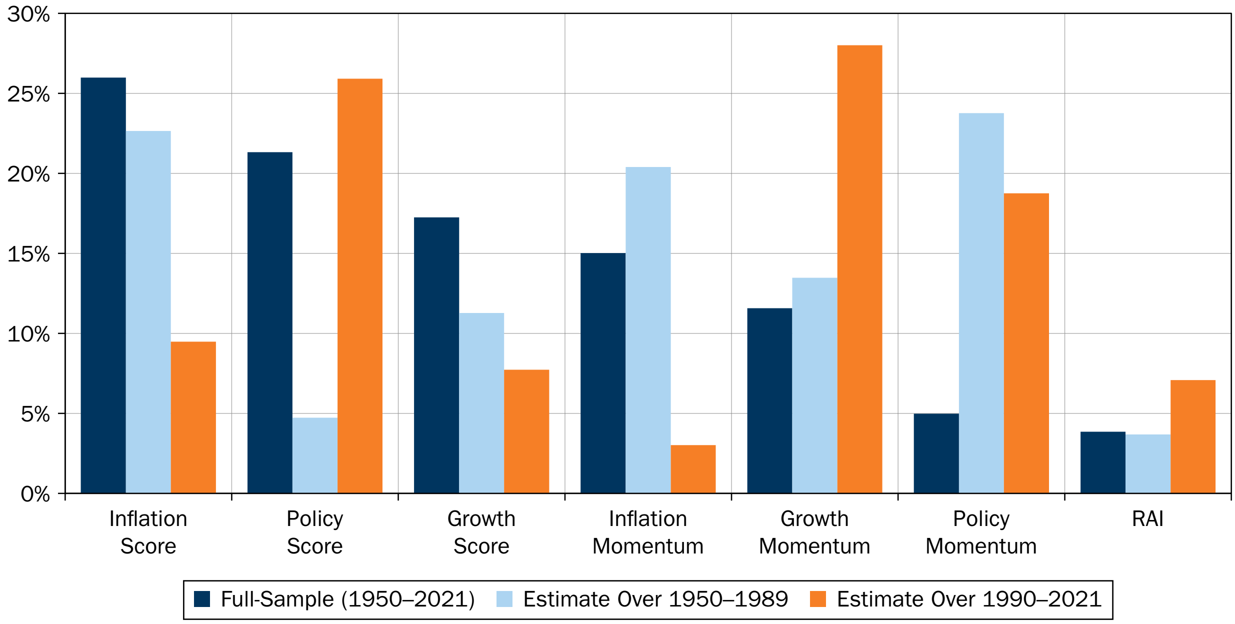
NOTES: The dotted line denotes unconditional probability. The bar chart shows the Shapley decomposition of the current probability (latest score in brackets). Data are for January 1950–November 2023.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

versus 60/40. During the COVID-19 recovery, the performance of gold was less consistent, but up until early 2023, the S&P GSCI has been the strongest asset class. The model struggled to capture the 2005–2007 oil bull market, which resulted from supply shortages following structural underinvestment. As a result, inflation picked up, with the growth score already declining going into the GFC. On the back of inflation

EXHIBIT 14

Importance of Cycle Scores in Determining Forward Equity vs. Bond Return < –10%

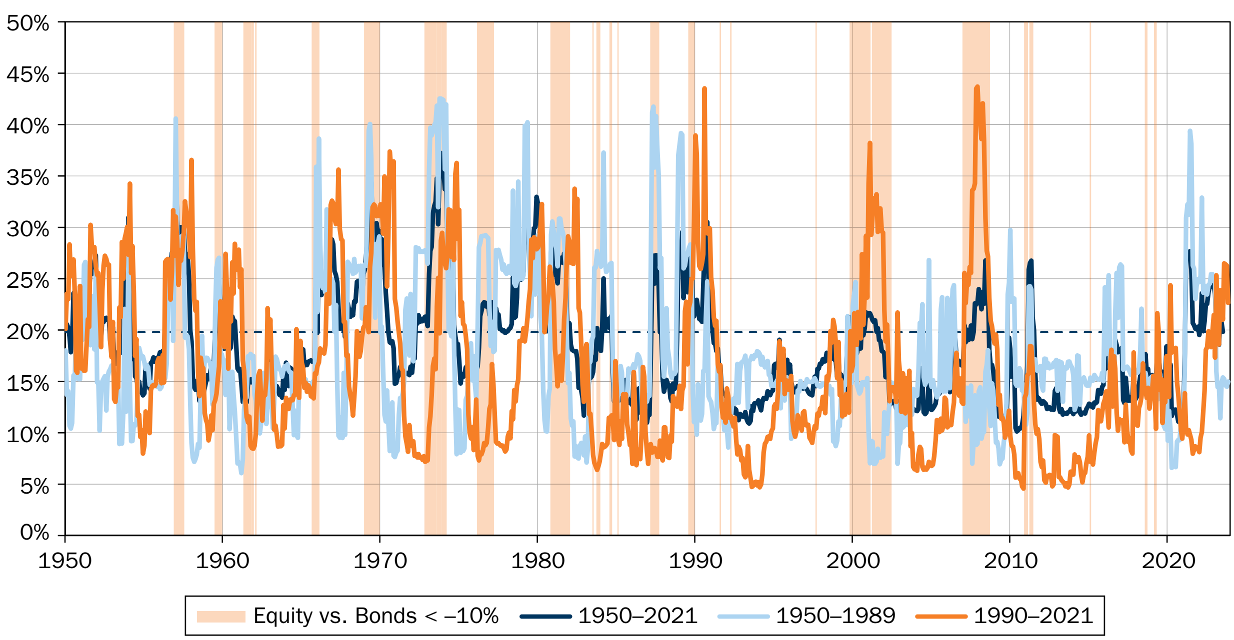


NOTES: Relative importance adds to 100%. The model is fitted on a January 1950–December 2021 sample.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

EXHIBIT 15

Random Forest Probability Estimates of Equity vs. Bond Drawdown Risk for Different Samples

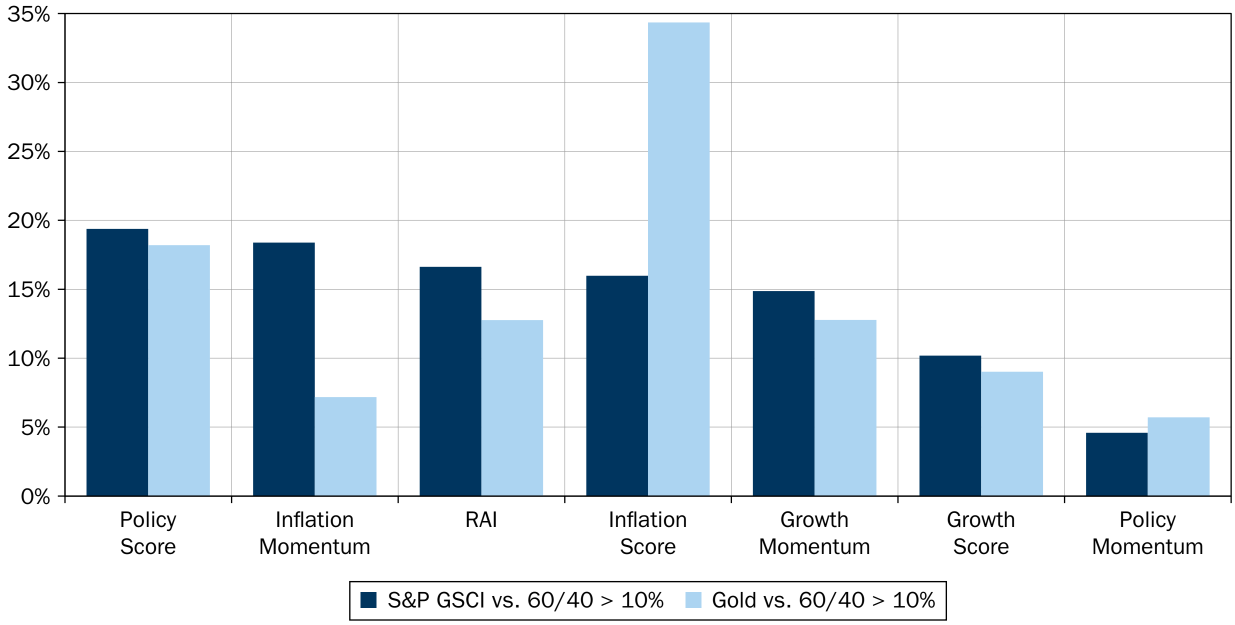


NOTES: The dotted line denotes unconditional probability. Data for January 1950–November 2023.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

EXHIBIT 16

Importance of Cycle Scores in Determining Forward Commodity Returns vs. 60/40 Portfolios



NOTES: Relative importance adds to 100%. The model is fitted on a January 1950–December 2021 sample.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

normalizing, the likelihood of commodity outperformance versus 60/40 has declined, consistent with weaker performance in 2023.

DYNAMIC ASSET ALLOCATION OVERLAYS BASED ON MACRO REGIMES

To benefit from changing macro regimes in real time, we build dynamic asset allocation (DAA) strategies based on their probabilities. The DAA overlays scale in and out of assets on a monthly basis, based on the regime probability (i.e., an allocation is increased linearly as the probability goes from the unconditional level to its 90th percentile). The aim is less to reflect short-term tactical views but rather to capture medium-term signals (12-month+ horizon) that impact forward return distributions. If the identified macro regimes have predictive power for market outcomes, it should be possible to capture some cyclical variation in cross-asset performance to add value over static weights in asset allocation. Results are before transaction costs, even though portfolio turnover is not very large 'on average' as the strategies focus on relatively rare tail outcomes.

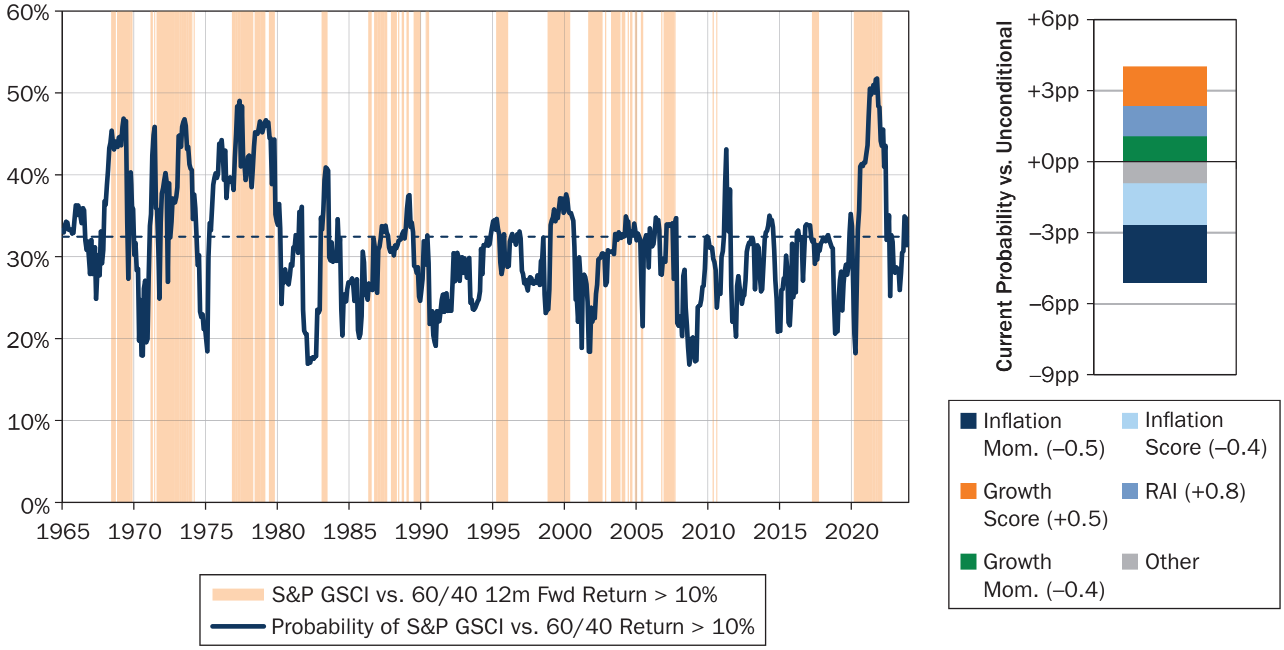
The first DAA overlay attempts to manage 60/40 tail risk by timing large drawdowns and recoveries. It scales into T-bills as the probability of a 60/40 drawdown in the next 12 months is rising and increases the exposure to the 60/40 portfolio to 150% using leverage (financed by T-bills) if there is elevated right-tail risk; Exhibit 18 illustrates the portfolio weights over time. The strategy captures the largest 60/40 drawdowns during inflationary times, such as the 1970s and COVID-19, as well as some of the recoveries.

The largest performance contribution from shifting to cash has come from avoiding the 60/40 drawdowns during the 1970s and recently, as shown in Exhibit 19.

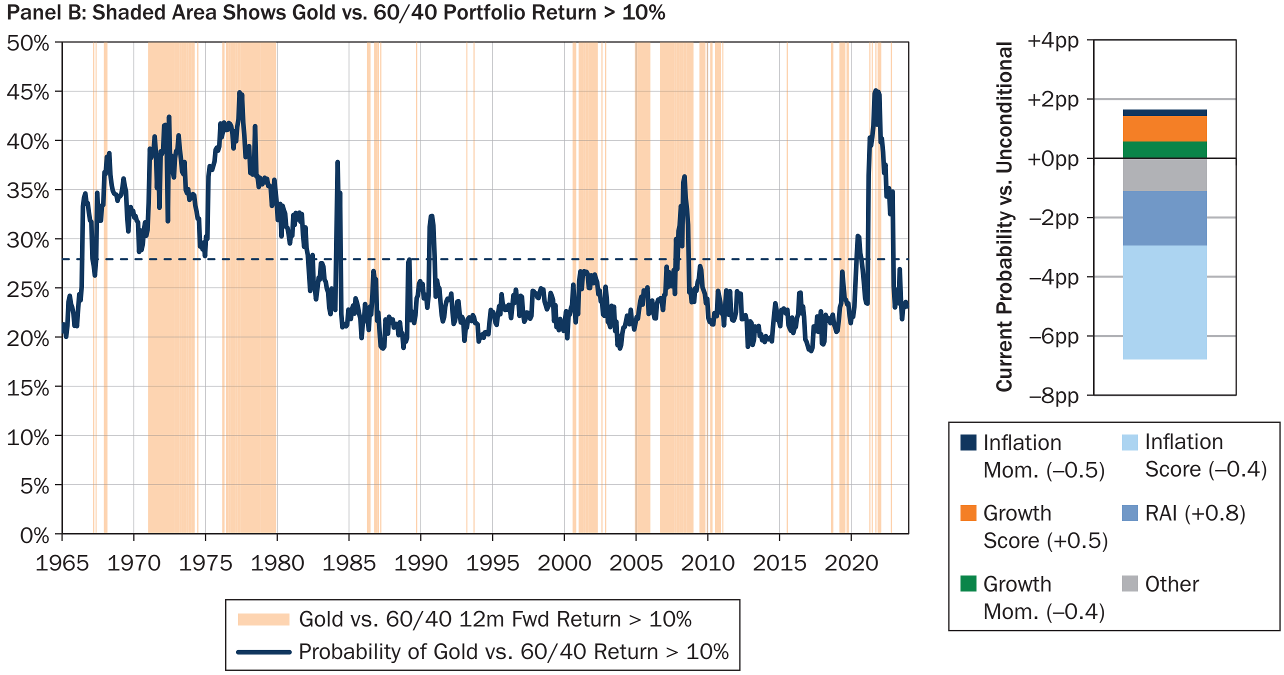
EXHIBIT 17

Random Forest Probability Estimates

Panel A: Shaded Area Shows S&P GSCI vs. 60/40 Portfolio Return > 10%



Panel B: Shaded Area Shows Gold vs. 60/40 Portfolio Return > 10%



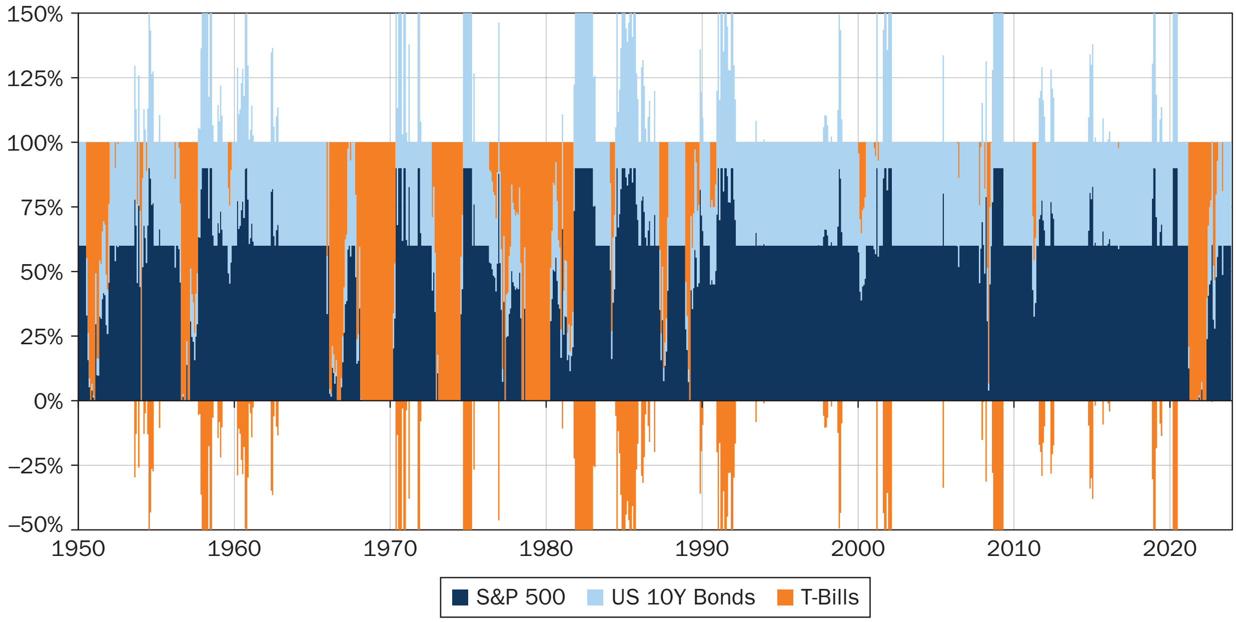
NOTES: The dotted line denotes unconditional probability. The bar chart shows the Shapley decomposition of the current probability (latest score in brackets). Data are for January 1965–November 2023.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

Since the 1990s, there has been less value added from being bearish, in part because 60/40 portfolios mostly performed very well. The 60/40 drawdown in 2022 was captured, but the model shifted bearish too early, in part because there was a longer than normal gap between inflation picking up in 2021 and policy tightening

EXHIBIT 18

DAA Portfolio Weights

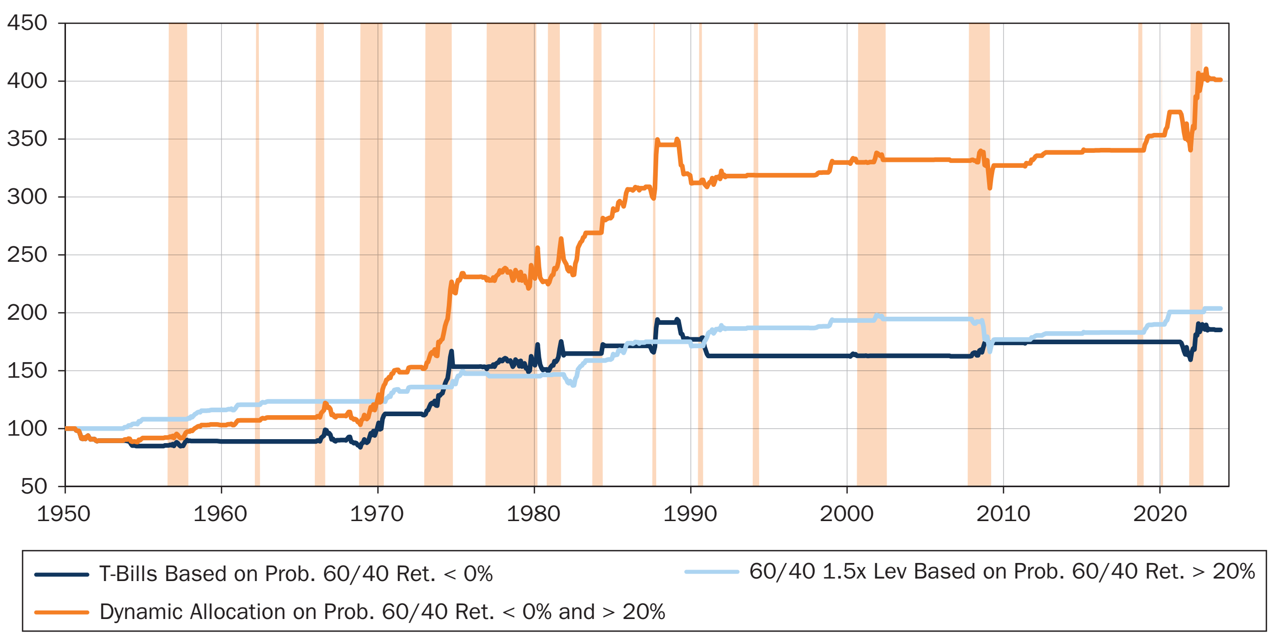


NOTE: A negative weight for T-bills denotes the use of leverage.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

EXHIBIT 19

Cumulative Relative Total Return Performance of DAA Strategy vs. Static US 60/40 Portfolio



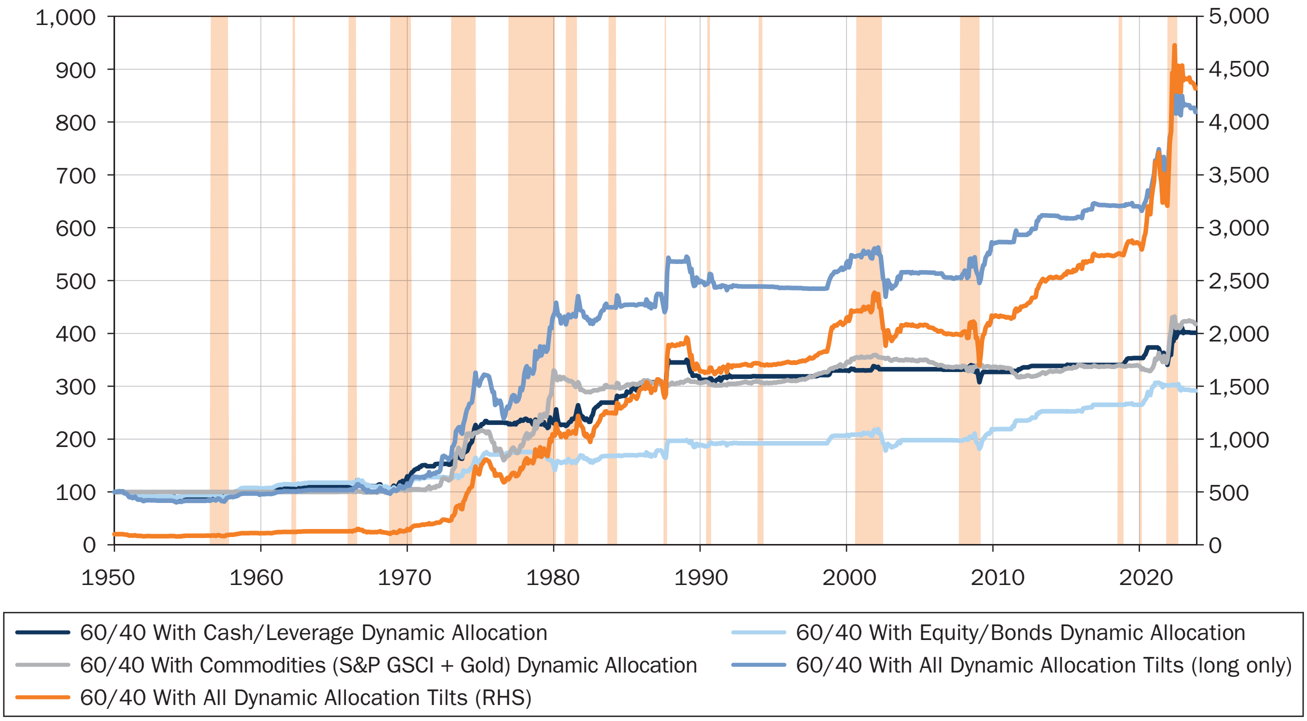
NOTE: Orange shading denotes 60/40 drawdowns.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

EXHIBIT 20

Cumulative Total Return and Performance of Dynamic Asset Allocation Strategies vs. Static US 60/40 Portfolio

Panel A: Cumulative Total Return



Panel B: Performance as of November 2023

		60/40 with Dynamic Asset Allocation Overlays							
	60/40 Portfolio	+ Cash	+ Cash/ Leverage	+ Equity/ Bonds	+ S&P GSCI	+ Gold	+ S&P GSCI/ Gold	+ All Dynamic Tilts	+ All Dynamic Tilts (long only)
Start Year	1950	1950	1950	1950	1965	1965	1950	1950	1950
Annualized Return	9%	10%	11%	11%	10%	11%	11%	15%	12%
Return p.a. vs. 60/40		0.9%	2.1%	1.6%	1.2%	1.6%	2.1%	5.7%	3.2%
Volatility	9%	8%	9%	10%	9%	12%	10%	12%	9%
1M 5% CVaR	–5%	–4%	–5%	–6%	–5%	–6%	–5%	–6%	–5%
TEV vs. 60/40		4%	5%	5%	7%	11%	5%	10%	7%
Sharpe Ratio	0.5	0.8	0.8	0.7	0.7	0.5	0.7	0.9	0.9
Return/CVaR	1.7	2.4	2.4	1.9	2.2	1.8	2.1	2.5	2.5
Information Ratio		0.2	0.5	0.3	0.2	0.1	0.4	0.6	0.5
					Average Allocation				
S&P 500	60%	49%	53%	56%	35%	33%	56%	45%	48%
US 10-Year Bonds	40%	33%	36%	44%	41%	37%	36%	36%	32%
T-Bills	0%	19%	11%	0%	7%	8%	1%	12%	13%
S&P GSCI	0%	0%	0%	0%	17%	0%	3%	3%	4%
Gold	0%	0%	0%	0%	0%	22%	4%	4%	6%

NOTES: In Panel A, "All dynamic allocation tilts," shown on the right-hand axis, sums the tilts from the cash/leverage dynamic allocation, the equity/bond dynamic allocation, and the commodity dynamic allocation (i.e., it is made up of 1.0× the cash/leverage tilt + 1.0 × the equity/bond tilt + 0.5× the GSCI tilt + 0.5× the gold tilt). "All dynamic allocation tilts (long only)" is the same but without short selling and leverage; it is constructed from the weights of the main strategy by removing the short positions and rescaling the long ones to sum to 100%. The orange shading denotes 60/40 drawdown.

SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

in 2022. The COVID-19 recovery was captured well, but during the GFC, the model scaled back into a leveraged 60/40 portfolio too early.

The second DAA overlay allocates between equities and bonds within a long-only portfolio based on the prospects for relative performance in the next 12 months. This focuses more on the source of the risk in the portfolio, which is usually equities, and is more relevant for investors who cannot have large cash allocations.

The last DAA overlays are to commodities, which tend to be more volatile than a 60/40 portfolio. To keep portfolio risk stable over time, we first create benchmark portfolios including commodities, targeting a volatility of 10% since 1965. We create the optimal portfolios for the respective macro regimes since 1965—for example, when the S&P GSCI outperformed a 60/40 portfolio materially—again targeting 10% volatility.[4] We then allocate across those portfolios based on the regime probabilities.

Exhibit 20 shows the cumulative performance and related stats. Our regime-based DAA strategies would have enhanced risk-adjusted returns relative to buy-and-hold 60/40 strategies. A combined strategy with all DAA overlays with the ability to go short and use leverage would have added almost 600 basis points (bps) of outperformance since 1950; even a long-only strategy with all tilts combined would have added about 300 bps. Avoiding or diversifying the inflationary 60/40 drawdowns of the 1970s and the COVID-19 recovery added material performance. That said, signals since the 1990s have been weaker, in part owing to lack of inflation risk. Especially during the Tech Bubble and GFC, the DAA overlays would have added risk too early.

CONCLUSION

Results from using a random forest algorithm for dynamic asset allocation are encouraging: Random forest probabilities based on our business cycle and risk appetite indicators captured several of the large tails around a US 60/40 portfolio, and related dynamic asset allocation overlays would have materially improved risk-adjusted returns since 1950. The approach is appealing because it is “supervised”—that is, input and output are controlled—and because it can capture nonlinearities and interaction effects in the data. Coupled with different visualizations such as Student trees and Shapley decompositions, it is well suited for more use in active asset management, which tends to be reluctant to use “black box” models to make investment decisions. Of course, as with most business-cycle investing frameworks, results are mostly in sample and may not capture all possible macro regimes; but results are relatively robust to out-of-sample testing.

High-inflation regimes have been more persistent and easier to capture, whereas growth cycles have been more difficult to capture, in part because markets lead growth and because of the potential for sharp reversals around inflection points. As a result, the value-added of the dynamic asset allocation overlays has been smaller since the 1990s and up until the COVID-19 crisis. Training a random forest on data since 1990 would have only better captured growth cycles since then (but would have struggled with inflationary cycles). This illustrates that structural changes in the

[4] Portfolios for S&P GSCI DAA overlays: *Benchmark optimal portfolio:* 40% S&P 500, 49% US 10-year bonds, 11% S&P GSCI, 0% T-Bills. *Overweight GSCI optimal portfolio:* 33% S&P GSCI, 67% T-Bills. *Underweight GSCI optimal portfolio:* 45% S&P 500, 48% US 10-year bonds, 7% T-Bills. Portfolios for Gold DAA overlays: *Benchmark optimal portfolio:* 36% S&P 500, 37% US 10-year bonds, 14% Gold, 13% T-Bills. *Overweight Gold optimal portfolio:* 33% Gold, 67% T-Bills. *Underweight Gold optimal portfolio:* 45% S&P 500, 48% US 10-year bonds, 7% T-Bills.

sample period will matter. A shorter-term market-timing approach, such as focusing on 3–6 month returns and incorporating higher-frequency macro and market data, could help better capture reversals, and incorporating structural regime breaks would likely enhance predictive power. Finally, incorporating forward views on macro variables could help better capture regimes and related cross-asset performance.

APPENDIX

DETAILS ON BUSINESS-CYCLE AND RISK APPETITE INDICATORS

The business-cycle indicators are based on a combination of macro and market data. We select four from each for growth, inflation, and policy. The variables are selected based on both data availability and frequency, historical consistency in capturing business cycles, and the lowest lag in data reporting. We calculate expanding *z*-scores (using data as they become available, at least a 10-year history is required) for the level score and one-year rolling *z*-score for momentum.

Growth

The macro variables aim to capture the growth cycle position with low unemployment versus NAIRU (nonaccelerating inflation rate of unemployment), positive output gaps, negative private sector financial balances, and high corporate profit margins indicating a late-cycle position. We also include cyclical risk premia; these usually increase materially during recessions and usually are declining when moving late cycle.

Inflation

We combine reported measures such as US CPI, core CPI, PPI, and PCE with market variables such as commodity prices (WTI and CRB) and short-term and long-term breakeven inflation (2 year and 10 year).[5] As a result, we capture both actual inflation and inflation expectations; during the 1970s, inflation expectations became unanchored, resulting in a much worse inflationary regime.

Policy

To capture monetary policy—central bank rate cycles, QE (quantitative easing), and credit conditions—we include the US federal funds rate versus its 10-year average, the year-over-year change in M2 money supply, and US lending standards. To capture fiscal policy, we include the fiscal deficit (as a percentage of GDP). To track the central bank policy cycle, we include market-implied changes to the federal funds rate, the two-year real rate, and the steepness of the US yield curve (2s10s); and more broadly, we include our US Financial Conditions Index (FCI).

Risk Appetite Indicator

The *risk appetite indicator* is a one-year rolling *z*-score of the total return performance of the S&P 500 versus US 10-year bonds and US cyclical versus defensive equities (based on sector classifications), as well as USD BAA credit spreads.

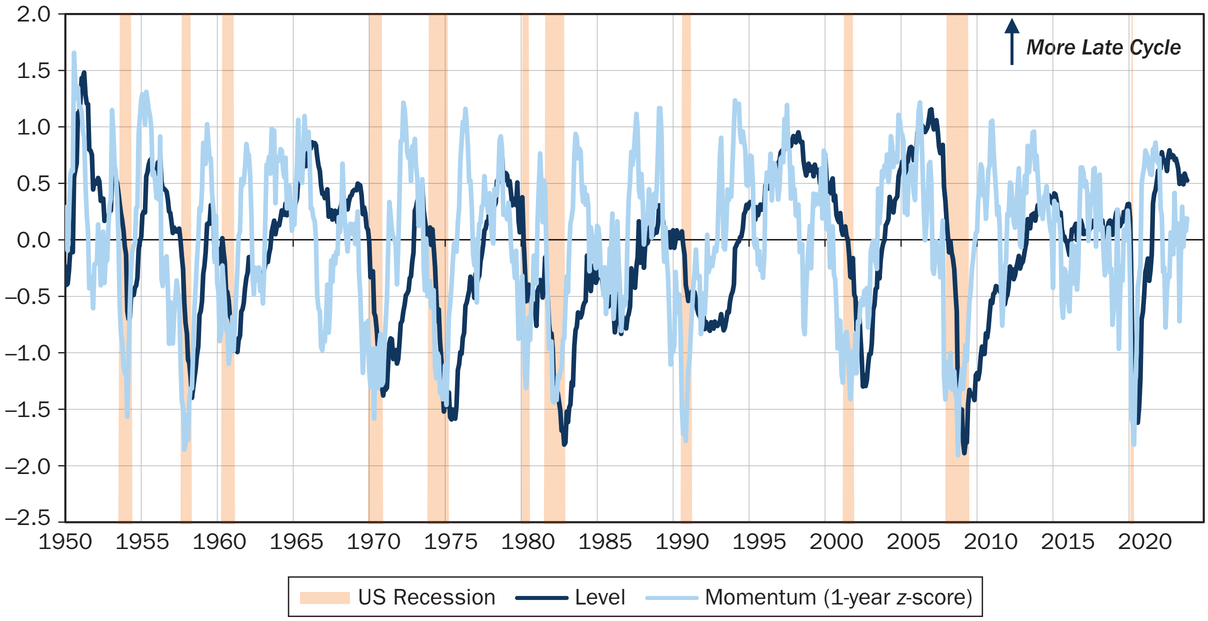
Exhibits A1–A3 show the level (expanding *z*-scores) and momentum (one-year *z*-score) scores for growth, inflation, and policy. Exhibit A4 shows the risk appetite indicator (RAI).

---

[5] For long-term US breakeven inflation and real yield data, we use a backcast based on a combination of economic and financial variables; see *Goldman Sachs: Global Rates Notes: Introducing a Backcast History of Traded Inflation*, August 26, 2020.

EXHIBIT A1

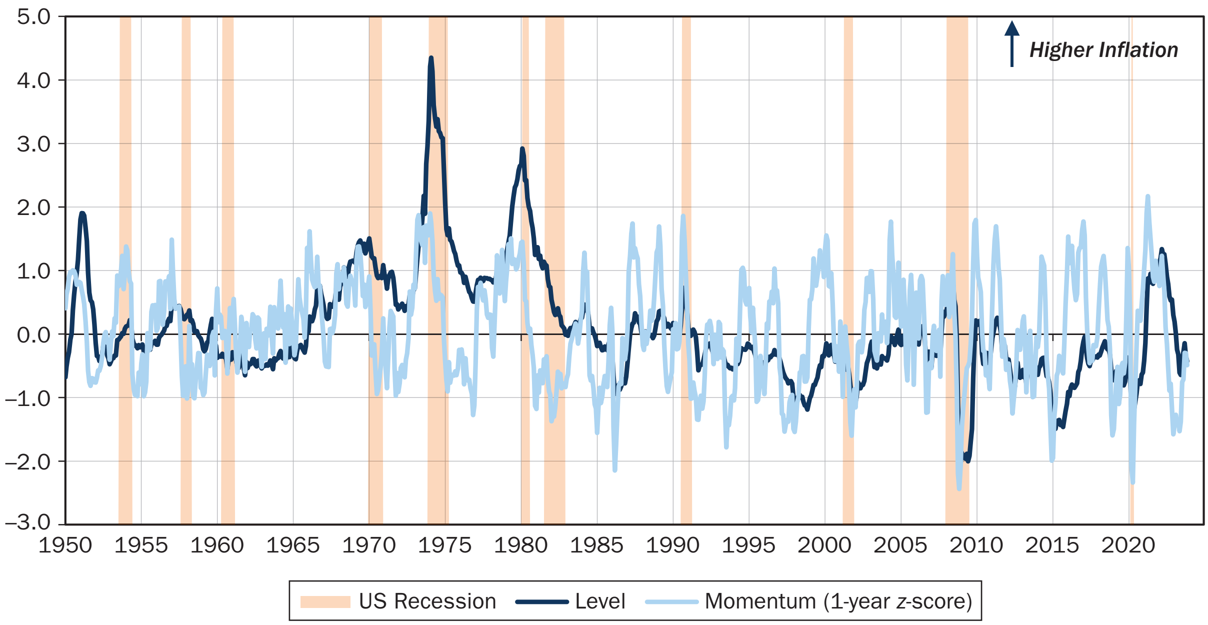
Growth Scores (level and momentum)



SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

EXHIBIT A2

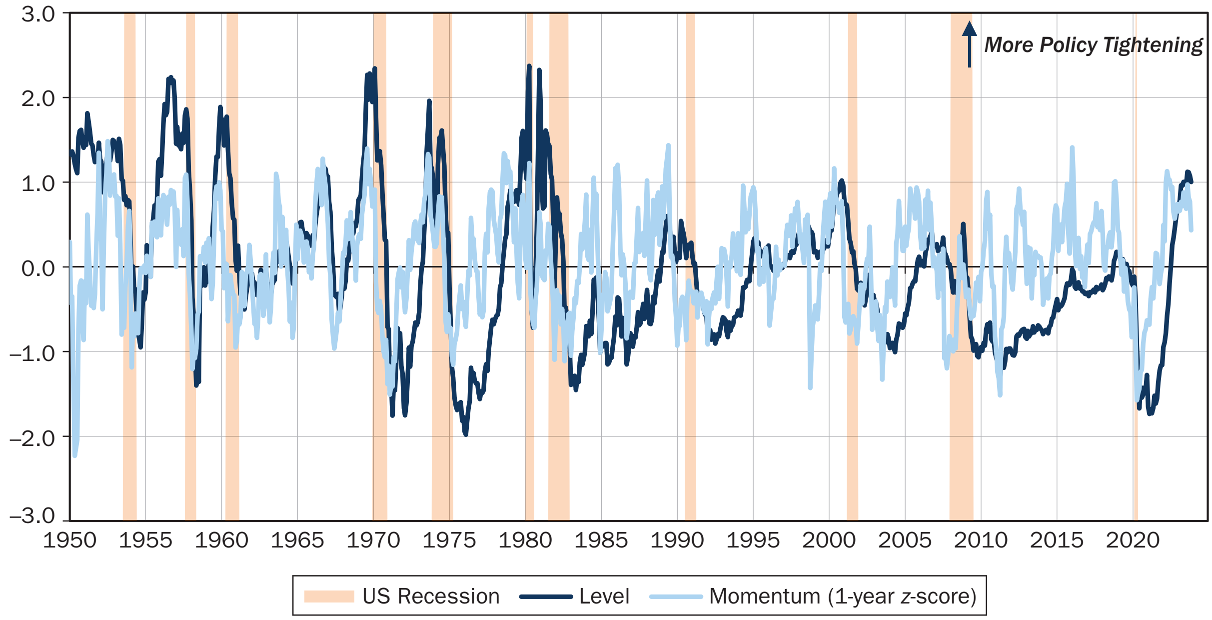
Inflation Scores (level and momentum)



SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

EXHIBIT A3

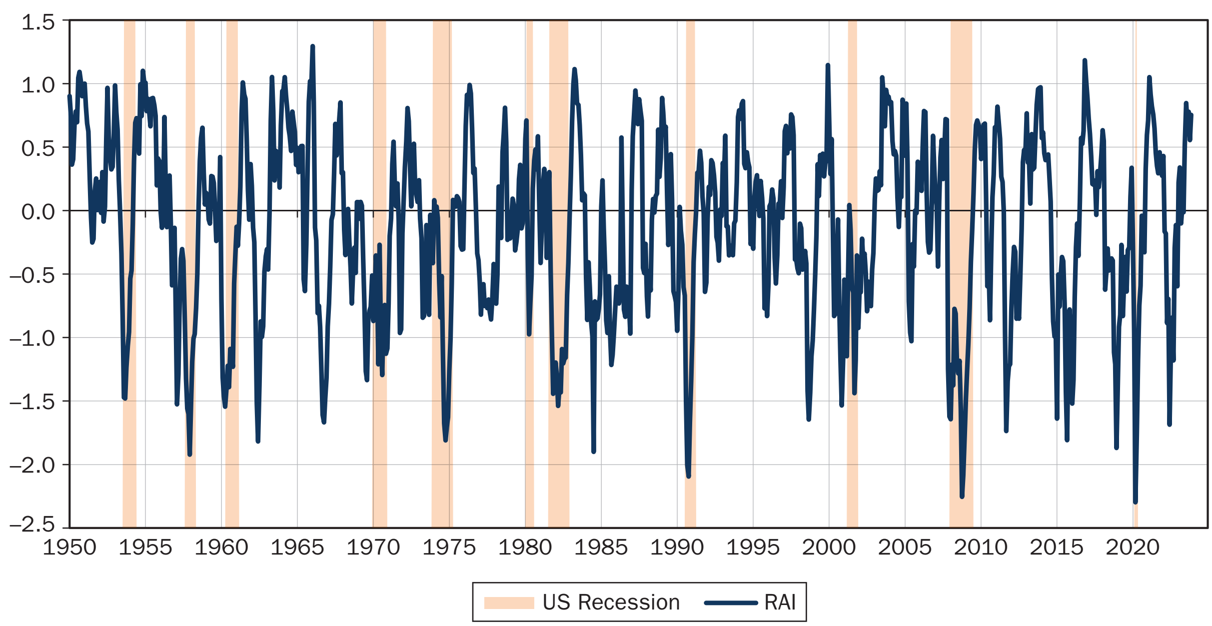
Policy Scores (level and momentum)



SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

EXHIBIT A4

Risk Appetite Indicator



SOURCES: Based on data from Haver Analytics, Bloomberg, Datastream, and Goldman Sachs Global Investment Research.

ACKNOWLEDGMENTS

We would like to thank Cecilia Mariotti, Peter Oppenheimer, Sharon Bell, Ingrid Tierens, Guillaume Jaisson, Tridibesh Dutta, Sid Bhushan, and Loretta Sunnucks for their helpful comments.

REFERENCES

Ang, A., and G. Bekaert. 2004. "How Regimes Affect Asset Allocation." *Financial Analysts Journal* 60 (2): 86–99.

Bartram, S. M., J. Branke, and M. Motahari. 2020. "Artificial Intelligence in Asset Management." *CFA Institute Research Foundation Literature Reviews* (August 28). https://www.cfainstitute.org/en/research/foundation/2020/rflr-artificial-intelligence-in-asset-management.

Blitz, D., T. Hoogteijling, H. Lohre, and P. Messow. 2023. "How Can Machine Learning Advance Quantitative Asset Management?" *The Journal of Portfolio Management* 49 (9): 78–95. https://www.pm-research.com/content/iijpormgmt/early/2023/01/11/jpm20231460.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Elkan, C. 2001. "The Foundations of Cost-Sensitive Learning." *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence* 2 (IJCAI'01): 973–978.

Géron, A. 2022. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd edition. Sebastopol: O'Reilly Media.

Ho, T. K. 1995. "Random Decision Forests." *Proceedings of 3rd International Conference on Document Analysis and Recognition* 1: 278–282.

Lee, Y., J. R. J. Thompson, J. H. Kim, W. C. Kim, and F. A. Fabozzi. 2023. "An Overview of Machine Learning for Asset Management." *The Journal of Portfolio Management* 49 (10): 31–63. https://www.pm-research.com/content/iijpormgmt/early/2023/08/12/jpm20231526.

Lundberg, S. M., and S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*: 4768–4777.

Nystrup, P., B. W. Hansen, H. Madsen, and E. Lindström. 2015. "Regime-Based versus Static Asset Allocation: Letting the Data Speak." *The Journal of Portfolio Management* 42 (1): 103–109.

Nystrup, P., B. W. Hansen, H. O. Larsen, H. Madsen, and E. Lindström. 2017. "Dynamic Allocation or Diversification: A Regime-Based Approach to Multiple Assets." *The Journal of Portfolio Management* 44 (2): 62–73.

Uysal, A. S., and J. M. Mulvey. 2021. "A Machine Learning Approach in Regime-Switching Risk Parity Portfolios." *The Journal of Financial Data Science* 3 (2): 87–108.

Sheikh, A. Z., and J. Sun. 2012. "Regime Change: Implications of Macroeconomic Shifts on Asset Class and Portfolio Performance." *The Journal of Investing* 21 (3): 36–54.

Yazdani, A. 2020. "Machine Learning Prediction of Recessions: An Imbalanced Classification Approach." *The Journal of Financial Data Science* 2 (4): 21–32.

Zhou, Y., Z. Zhou, and G. Hooker. 2018. "Approximation Trees: Statistical Stability in Model Distillation." *arXiv*:1808.07573.