Leading Indicators of S&P 500 Drawdowns:

A Horizon-Specific Macro-Financial Framework

Alexander Uyar

March 20, 2026

Abstract

This paper presents a comprehensive empirical study of 30 macroeconomic and financial indicators as predictors of S&P 500 equity-market drawdowns across four forecast horizons (1, 3, 6, and 12 months), using univariate probit models, ridge-penalised logistic regression, and OLS models. The standout short-horizon performers are the National Financial Conditions Index (AUC=0.725 at 1M), and VIX (AUC=0.721 at 1M), with HY Option-Adjusted Spread close behind (AUC=0.710 at 1M). At the 3-month horizon, HY OAS leads (AUC=0.729), followed by ISM Services Employment (AUC=0.725). At the medium term (6M), NFCI (AUC=0.719) maintains its predictive strength, while ISM Services Prices Paid emerges as a powerful medium-term signal (AUC=0.633). Over 12 months, VIX (AUC=0.696) and M2 YoY Growth (AUC=0.695) lead the field. For the multi-factor analysis, we construct four separate ridge-penalised logistic regression models, one per forecast horizon, each selecting the best group representative at that specific horizon. Out-of-sample performance is evaluated via an expanding-window walk-forward procedure, producing approximately 300 OOS observations per model. The four models achieve OOS AUCs of 0.600 (1M), 0.694 (3M), 0.737 (6M), and 0.684 (12M), substantially exceeding the benchmark (AUC=0.5). We additionally present a Current Market Assessment section that applies the full-sample fitted models to the most recent data releases, reporting estimated drawdown probabilities with 95% delta method confidence intervals, and a Historical Drawdown Episode Analysis evaluating model performance across nine major market drawdowns since 1990. Limitations including look-ahead bias in indicator construction, multicollinearity, and the inherent unpredictability of sentiment-driven crises are discussed explicitly in a dedicated section.

1 Introduction

The prediction and management of equity-market drawdowns is one of the central challenges facing portfolio managers. While a large literature has examined individual indicators - from the yield curve (Estrella & Hardouvelis, 1991) to credit spreads (Gilchrist & Zakrajsek, 2012) and valuation ratios (Shiller, 2000), comparatively few studies systematically evaluate a broad set of indicators within a common methodological framework and across multiple forecast horizons. This paper asks a focused empirical question: do broad macroeconomic and financial indicators carry statistically reliable signals of future S&P 500 drawdowns at horizons ranging from one month to twelve months? Rather than testing a single hypothesis, we adopt a comprehensive screening approach across 30 diverse indicators, allowing the data to reveal which combinations of signals, across which horizons, generate the strongest predictive power.

We adopt a peak-to-trough drawdown definition: for each forecast origin (t) and horizon (h), the drawdown indicator is triggered when the highest price within the forward window from t+1 to t+h is followed by a decline of 10% or more to the subsequent trough. This isolates genuine price reversals rather than declines from an already-depressed level, and is consistent with how drawdown risk is described in practice. We also estimate four separate horizon-specific models (1M, 3M, 6M, 12M), each selecting its own optimal group representative by AUC, allowing the model architecture to adapt to the differing economic transmission lags at each horizon. Out-of-sample performance is evaluated using a walk-forward procedure: the model is re-fitted each month using all available data up to that point, with predictions made one step ahead. This generates approximately 300 genuine out-of-sample predictions per model, covering the period from January 2000 onwards.

The paper proceeds as follows. Sections 3–11 present univariate results by indicator group. Sections 12–15 cover cross-indicator comparison, modelling challenges, model specifications, and full results. Sections 16–17 analyse historical episodes and the current market assessment. Section 18 discusses limitations and Section 19 concludes.

2 Data and Methodology

2.1 Dependent Variable: Peak-to-Trough Drawdown

The primary dependent variable is a binary drawdown indicator derived from the S&P 500 Index (Bloomberg ticker: SPX). We define the forward peak-to-trough drawdown for month t and horizon h as follows.
Let $P_1, P_2, \ldots, P_h$ denote the S&P 500 monthly closing prices in the forward window $[t+1,\, t+h]$. We define the intra-window peak as:

$$P^* = \max_{i \in \{1,\ldots,h\}} P_i \quad \text{at position } k^*$$

and the subsequent trough as:

$$T^* = \min_{i \in \{k^*+1,\ldots,h\}} P_i$$

The peak-to-trough drawdown is then:

$$\mathrm{DD}(t,h) = \frac{P^* - T^*}{P^*} \times 100$$

The binary drawdown indicator equals one if DD(t,h) $\geq$ 10% (or $\geq$ 5% for the 1-month horizon) and zero otherwise.

Following the threshold convention established by Lleo and Ziemba (2012), we classify it as a drawdown event when it reaches or exceeds 10%. Unlike their definition — which measures the maximum decline from the signal date within one year — our peak-to-trough formulation exclusively captures episodes where a genuine intra-window price peak is followed by a material correction, rather than penalising declines from an already-depressed starting level. At a one-month horizon, a true peak-to-trough drawdown cannot be defined: with only a single forward month-end price, there is no meaningful "peak" and "trough" within the window. We therefore use a simpler one-month definition for h=1, measuring the month-to-month decline from t to t+1, and apply a lower threshold of 5% to capture economically meaningful short-horizon drops. For the longer horizons (h=3,6,12), we use the full peak-to-trough definition within the forward window and apply a 10% threshold. All models are estimated using data from January 1990 to December 2024. Data from 2025 are used exclusively for the out-of-sample 'current market assessment' and are not included in model estimation. Over the 1990–2024 sample, these definitions imply event base rates of 9.5% at 1M (40 events), 4.0% at 3M (17), 15.7% at 6M (66), and 20.7% at 12M (87).

Table 1 summarises the identified drawdown episodes at the 3-month horizon. Figure 1 shows the full time series of forward peak-to-trough drawdowns at all four horizons, with NBER recessions shaded in grey.

Table 1: Historical Drawdown Episodes Identified by the 3-Month Indicator (Peak-to-Trough ≥10%)

Signal Start	Signal End	Peak (SPX)	Trough Date	Drawdown (%)	Months Flagged
1990-05	1990-08	361	1990-10	15.8%	3
1998-05	1998-08	1,091	1998-08	12.2%	3
2000-08	2000-09	1,518	2001-02	18.3%	1
2000-12	2001-02	1,320	2001-03	12.1%	2
2001-06	2001-08	1,224	2001-09	15.0%	2
2002-03	2002-09	1,147	2002-09	28.9%	6
2007-10	2007-12	1,549	2008-03	14.6%	2
2008-07	2009-02	1,267	2009-01	34.8%	7
2010-03	2010-05	1,169	2010-06	11.9%	2
2011-06	2011-08	1,321	2011-09	14.3%	2
2018-09	2018-10	2,914	2018-12	14.0%	1
2019-12	2020-03	3,231	2020-03	20.0%	3
2022-03	2022-04	4,530	2022-09	20.9%	1

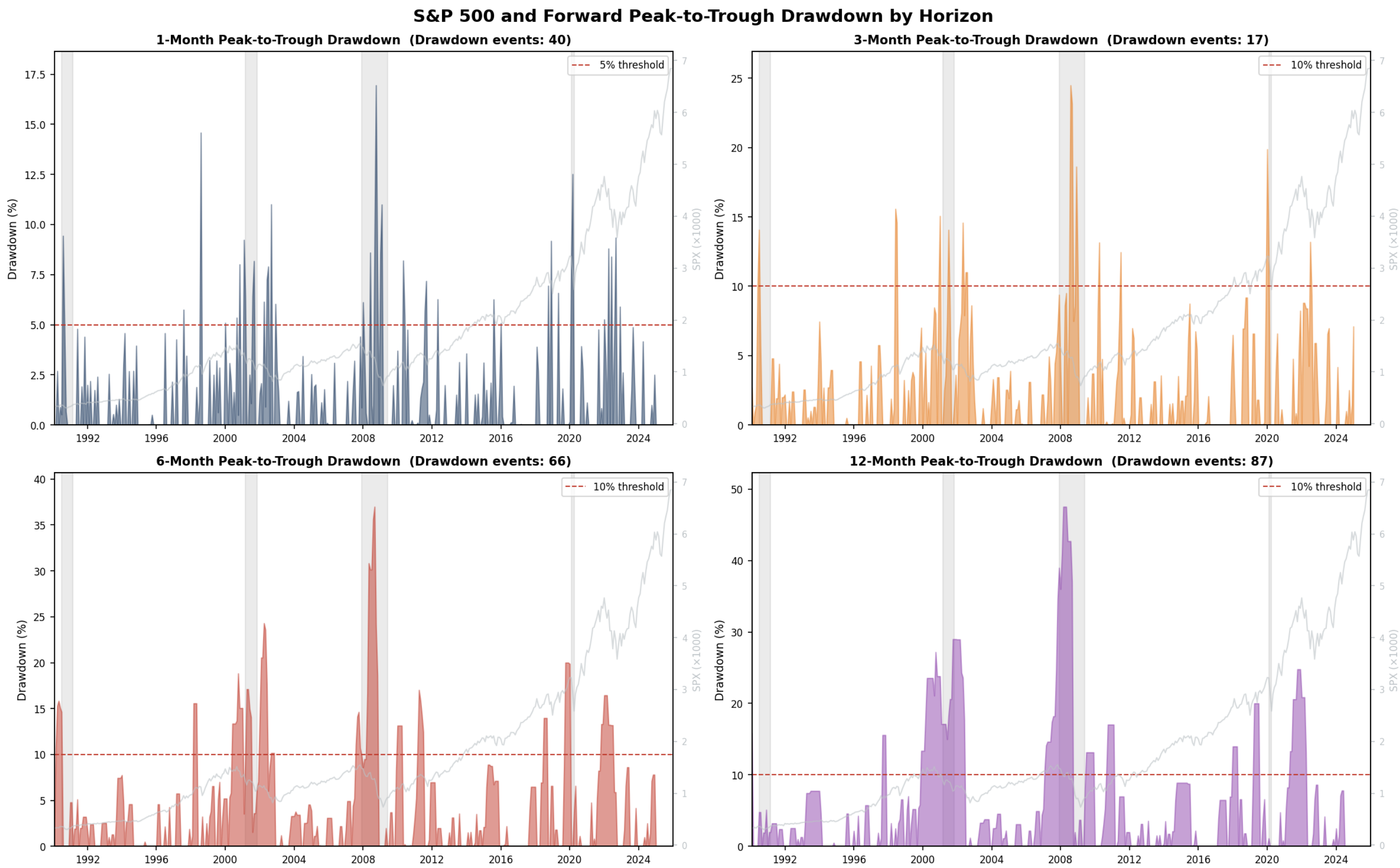

Figure 1: S&P 500 and Forward Peak-to-Trough Drawdown (full sample) by Horizon (h=1, 3, 6, 12 months). Coloured fill shows the realised peak-to-trough drawdown (%). Dashed red line indicates the classification threshold (5% for h=1M; 10% for h=3, 6, 12M). Grey shading indicates NBER recession periods.

2.2 Indicators

We test 30 leading indicators spanning nine thematic groups. Five of the nine groups are based on survey data (ISM Manufacturing, ISM Services) or financial market prices (Yield Curve, Credit, Volatility). The remaining four groups — Valuation, Financial Conditions, Money & Liquidity, and Consumer & Labour Market — reflect fundamental macroeconomic or composite conditions.

ISM Manufacturing indicators include the Composite PMI, New Orders, Production, Employment, Inventories, and Supplier Deliveries sub-indices. ISM Services indicators include the Composite PMI, New Orders, Business Activity, Employment, and Prices Paid sub-indices (available from July 1997). The Valuation group contains a single indicator, the Bond-Stock Earnings Yield Differential (BSEYD) (Lleo & Ziemba, 2012), derived from the reciprocal of the Shiller CAPE ratio (Shiller, 2000) minus the 10-year Treasury yield, standardised using a 60-month rolling z-score. The Yield Curve group contains the 10Y–3M and 10Y–2Y spreads. The Credit group contains the ICE BofA US High-Yield Option-Adjusted Spread. The Volatility group contains the CBOE VIX, ICE BofA MOVE Index, and CBOE SKEW Index. The Financial Conditions group contains the Bloomberg Financial Conditions Index (BFCI) and Chicago Fed National Financial Conditions Index (NFCI). The Money & Liquidity group contains M2 Level, M2 YoY Growth, a QE-adjusted M2 residual, and Federal Reserve Total Assets. The Consumer & Labour Market group contains the University of Michigan Consumer Sentiment Index, Conference Board Consumer Confidence Index, US Unemployment Rate, the Sahm Rule (a rate-of-change indicator derived from unemployment), Initial Jobless Claims, and US Housing Permits (Total).

2.3 Statistical Models

All indicators are z-score standardised using full-sample statistics prior to modelling. For the binary drawdown indicator, we estimate univariate probit models at each of the four forecast horizons, reporting the AUC (Area Under the ROC Curve) as the primary performance metric. For the continuous drawdown measure, we use ordinary least squares regression and report R^2. In the multi-factor stage, we use ridge-penalised logistic regression (penalty parameter $\lambda=1.0$) to reduce the effects of predictor collinearity and limit overfitting to the small number of drawdown events.

For out-of-sample evaluation, we use an expanding-window walk-forward procedure. Starting from January 2000 (requiring a minimum 120-month training window), we re-fit the ridge logit at each month using all data available up to that point, and predict the probability for that month. This produces approximately 300 genuine out-of-sample predictions per model — and this is the standard approach in the macro-finance forecasting literature (Goyal & Welch, 2008; Rapach & Zhou, 2013). AUC is computed on these OOS predictions only.

3 ISM Manufacturing Indicators

The Institute for Supply Management (ISM) Manufacturing Purchasing Managers' Index is one of the most closely watched leading indicators in macroeconomics. Released on the first business day of each month, the composite PMI and its sub-components provide a real-time pulse on manufacturing sector conditions. In general, readings above 50 signal expansion; below 50, contraction. In our sample of 420 monthly observations (January 1990 to December 2024), six manufacturing sub-indices are tested: Composite PMI, New Orders, Production, Employment, Inventories, and Supplier Deliveries.

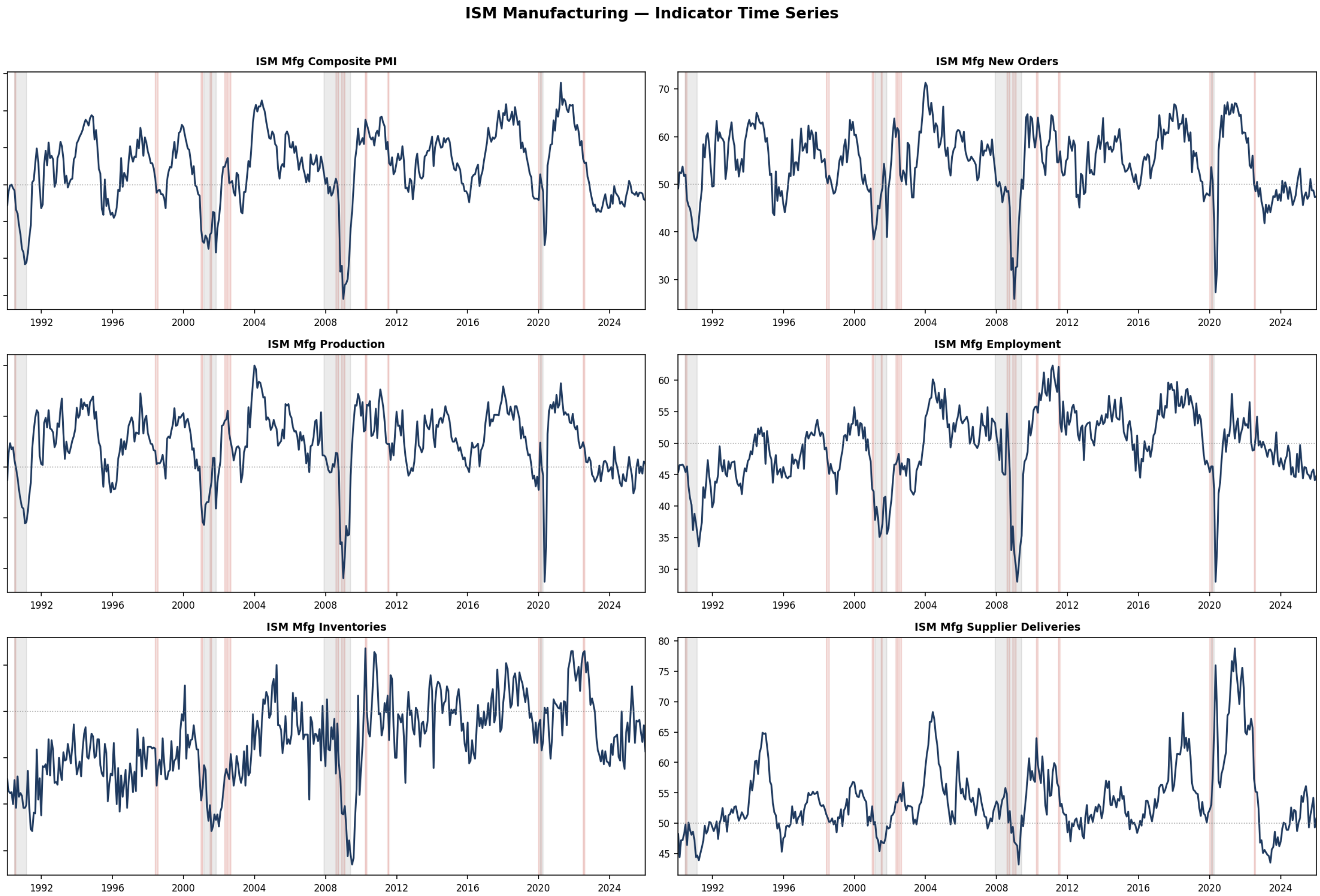

Figure 2: ISM Manufacturing Indicator Time Series. Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a $\geq$10% drawdown event). Grey shading indicates NBER recession periods. Horizontal dotted line at 50 marks the ISM expansion/contraction boundary.

3.1 Results

ISM Manufacturing New Orders is the strongest performer in this group across short horizons, achieving AUC=0.615 at 1M and AUC=0.692 at 3M, both statistically significant. ISM Mfg Production follows closely (AUC=0.612 at 1M, AUC=0.665 at 3M). ISM Mfg PMI achieves AUC=0.596 at 1M and 0.664 at 3M. ISM Mfg Employment is a marginally significant predictor at 1M (AUC=0.569, $p<0.10$) and 3M (AUC=0.646, $p<0.10$). ISM Mfg Inventories shows modest predictive power (AUC=0.618 at 3M), while ISM Mfg Supplier Deliveries is relatively weak at short horizons but improves materially at 6M (AUC=0.561, $p<0.10$) and 12M (AUC=0.574, $p<0.05$).

3.2 Economic Rationale

The strong performance of ISM Manufacturing at the 1–3 month horizons is consistent with its role as a real-time gauge of cyclical demand (Koenig, 2002; Lahiri & Monokroussos, 2013). Sharp declines in new orders signal an imminent slowdown in output and revenues, which tends to translate quickly into earnings revisions, risk-premium repricing, and therefore near-term equity weakness. Its weaker performance at 6–12 months is

also intuitive: manufacturing conditions are relatively volatile and often mean-revert within the year, so the information content of a single deterioration signal tends to decay as the forecast window lengthens.
Supplier Deliveries shows the opposite profile — less predictive at short horizons but more predictive at 12 months — because it proxies capacity constraints and supply-chain stress, which typically influence the economy with longer and more indirect lags. Persistently elevated delivery times can indicate late-cycle overheating, stretched inventories, and rising input pressures — conditions that can precede policy tightening, margin compression, and the broader slowdowns that drive larger equity drawdowns over longer horizons.

4 ISM Services Indicators

The ISM Services PMI has been published since July 1997, limiting this group to 330 monthly observations. Services account for approximately 77% of US GDP, making these indicators particularly relevant for equity market forecasting in the modern era. We test five sub-indices: Composite PMI, New Orders, Business Activity, Employment, and Prices Paid.

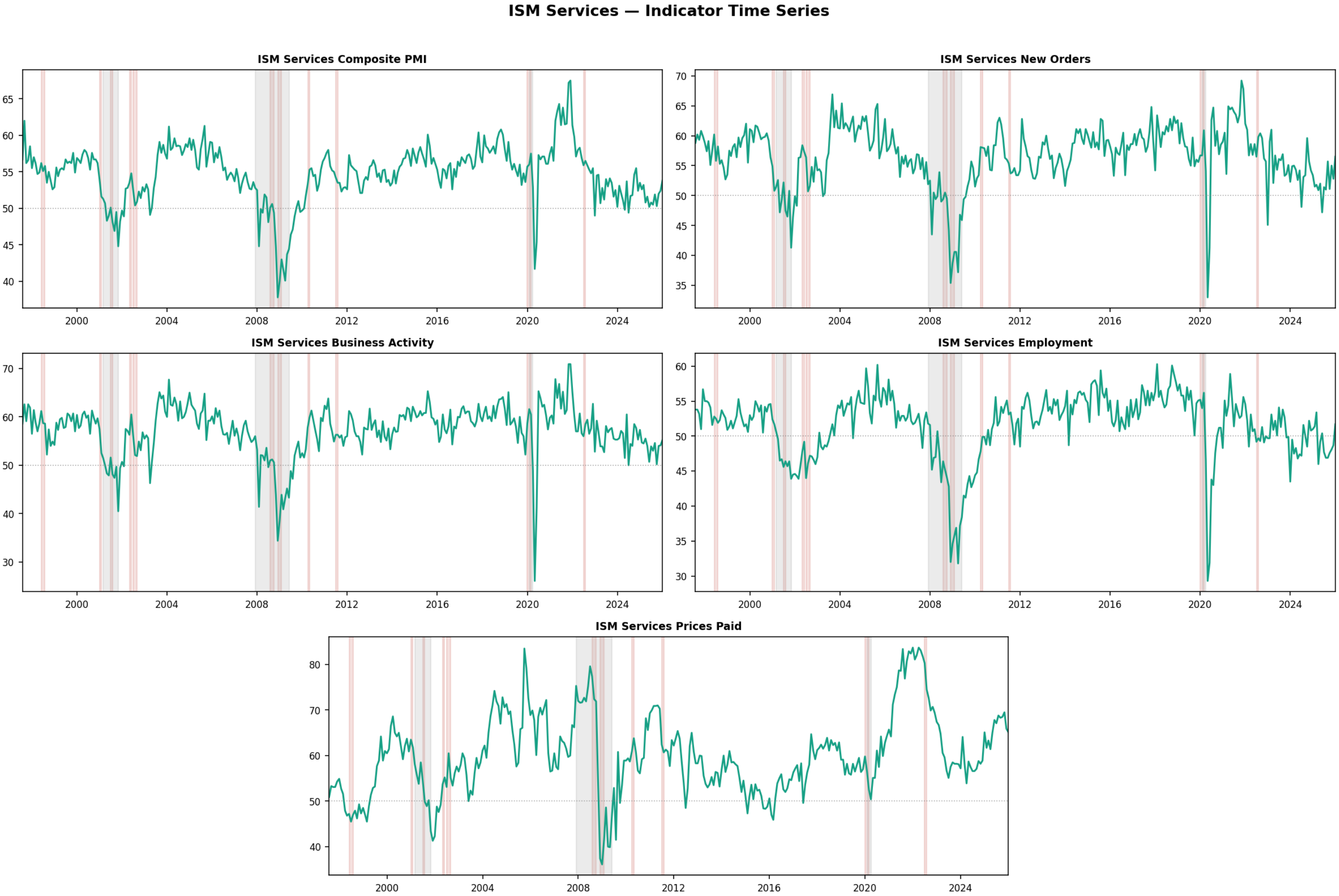

Figure 3: ISM Services Indicator Time Series. Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a ≥10% drawdown event). Grey shading indicates NBER recession periods.

4.1 Results

ISM Services Employment is the standout performer in this group, achieving AUC=0.598 at 1M ($p<0.10$), AUC=0.725 at 3M ($p<0.05$), AUC=0.629 at 6M ($p<0.05$), and AUC=0.614 at 12M ($p<0.05$) — the strongest sustained AUC profile among ISM indicators. ISM Services PMI reaches AUC=0.713 at 3M ($p<0.05$). ISM Services Business Activity achieves AUC=0.674 at 3M and AUC=0.631 at 6M. ISM Services New Orders produces AUC=0.695 at 3M. ISM Services Prices Paid is weak at short horizons but emerges as a powerful medium-term signal (AUC=0.633 at 6M, $p<0.05$; AUC=0.630 at 12M, $p<0.05$), ultimately selected as the best ISM Services representative in both the 6M and 12M multi-factor models.

4.2 Economic Rationale

ISM Services Employment tends to be most informative at the three-month horizon because hiring decisions in services typically respond to demand conditions with a short lag. Firms often wait for weaker activity to persist before adjusting headcount (Sahm, 2019), so employment indicators can emerge as a warning signal after the initial demand slowdown but before it shows up in broader earnings and risk pricing. A softening in services hiring therefore reflects a near-term demand reassessment that markets may not have fully incorporated, making it a useful early signal over the subsequent quarter. ISM Services Prices Paid displays a different timing, with predictive content concentrated at 6–12 months, consistent with the monetary policy channel. Persistent services-inflation pressure increases the likelihood that the Federal Reserve maintains a restrictive stance (or tightens further), which can weigh on equities via higher real discount rates and tighter financial conditions. Because this transmission typically operates with multi-quarter lags (Bernanke & Gertler, 1995), Prices Paid functions as a longer-horizon risk indicator rather than a short-horizon stress gauge. Finally, the comparatively strong performance of services indicators at the 3M horizon is consistent with the structure of the modern US economy and equity market: services dominate domestic activity, and S&P 500 earnings are heavily exposed to sectors such as technology, healthcare, and financial services. As a result, services survey signals can map more directly into near-term earnings expectations than purely manufacturing-focused measures.

5 Valuation: Bond-Stock Earnings Yield Differential (BSEYD)

5.1 Indicator Construction

The Bond-Stock Earnings Yield Differential (BSEYD), formalised as a drawdown predictor by Lleo and Ziemba (2012, 2015a, 2015b), is defined as the earnings yield of equities minus the risk-free bond yield:

$$\text{BSEYD}(t) = \frac{1}{\text{CAPE}(t)} - y_{10}(t)$$

where CAPE is the Shiller Cyclically Adjusted Price-Earnings ratio and y_{10} is the 10-year Treasury yield. A positive BSEYD indicates equities offer a yield premium over bonds (cheap relative to bonds); a negative BSEYD suggests equities are expensive. We standardise BSEYD using a 60-month rolling z-score to produce a stationary, regime-adjusted signal.

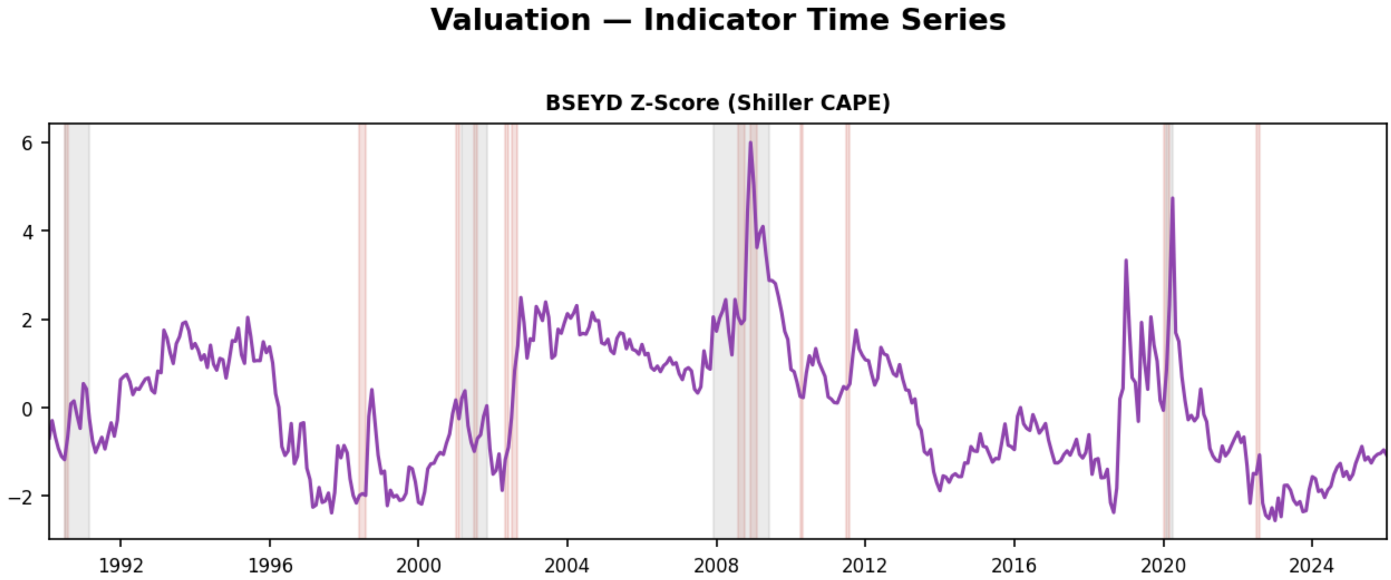

Figure 4: BSEYD Z-Score Time Series (Shiller CAPE). BSEYD standardised using a 60-month rolling z-score. Positive values indicate equities are cheap relative to bonds; negative values indicate expensive equities. Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a ≥10% drawdown event). Grey shading indicates NBER recession periods.

5.2 Results

The BSEYD produces only weak univariate predictive power with the peak-to-trough drawdown definition: AUC ranges from 0.509 to 0.542 across all four horizons with no statistically significant results. However, BSEYD proves highly significant in the multi-factor models, achieving t-statistics of −2.19 (p=0.028) in the 1M model and −2.93 (p=0.003) in the 6M model. The negative sign confirms that expensive equity valuations (low BSEYD) are associated with higher drawdown probability — consistent with its role as a valuation anchor that conditions the overall drawdown probability when combined with credit spreads and financial conditions. In the 3M model the t-statistic is −0.98 (p=0.329), reflecting partial absorption by correlated credit and volatility signals at that horizon. In the Current Market Assessment, BSEYD is consistently among the top three driver indicators.

5.3 Economic Rationale

Valuation is inherently a long-run mean-reversion concept: equities can remain expensive or cheap for extended periods before reverting to fundamental value. This explains why the BSEYD shows little univariate predictive power at any individual horizon — it has no fixed lead time. Its power emerges in the multi-factor context, where it acts as an "elevator" for conditional drawdown probabilities: when both credit spreads are wide AND equities are expensive (low BSEYD), the probability of a drawdown is materially elevated relative to periods of cheap equity valuations.

6 Yield Curve

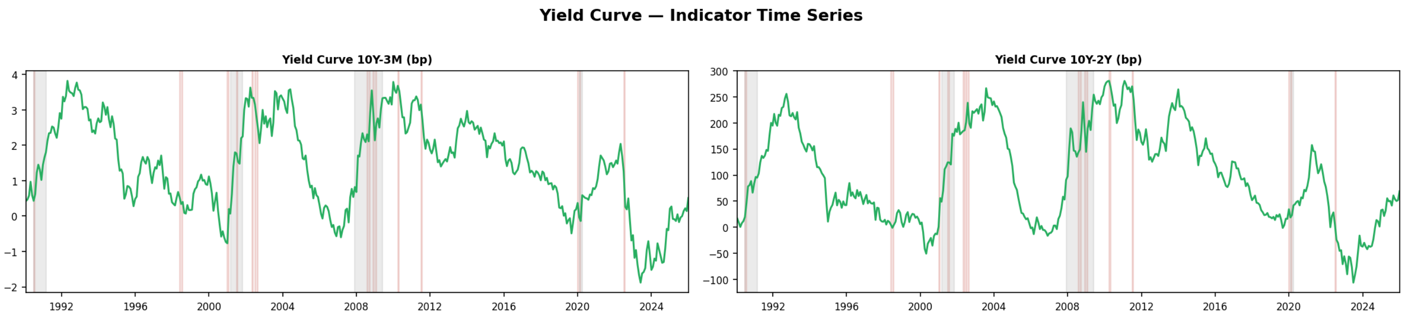

Figure 5: Yield Curve Indicator Time Series. 10Y–3M and 10Y–2Y Treasury yield spreads (basis points). Negative values indicate yield curve inversion. Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a ≥10% drawdown event). Grey shading indicates NBER recession periods.

6.1 Results

The 10Y–3M Treasury spread and 10Y–2Y spread exhibit modest univariate predictive power across all horizons: AUCs range from 0.504 to 0.554. These results are weaker than the historical reputation of yield curve inversion as a recession predictor would suggest. A key reason is that with the peak-to-trough definition, drawdown events are concentrated around three distinct episodes (2000–02, 2007–09, 2022), and the yield curve inversion of 2019–20 was followed by a very brief drawdown before rapid recovery, introducing signal noise.

6.2 Economic Rationale

The yield curve is one of the most studied recession predictors in macroeconomics, typically exhibiting lead times of 12–24 months. However, equity market corrections often precede the economic recession itself. This mismatch means that while the yield curve predicts NBER recessions reliably, it may fire too early or too late relative to the equity peak-to-trough episode. At the 12-month horizon, the 10Y–3M spread achieves AUC=0.550, reflecting the slow-moving nature of the monetary tightening transmission.

In the multi-factor models, the yield curve contribution emerges more clearly: the 10Y–3M spread is selected as the Yield Curve group representative for both the 1M and 3M models, and achieves t-statistics of 1.02 (1M) and 1.05 (3M). The 10Y–2Y spread is selected for the 6M model, achieving a significant t-statistic of 2.39 (p=0.017), consistent with the 6-month transmission lag from monetary tightening to corporate credit conditions.

7 Credit: High-Yield Option-Adjusted Spread

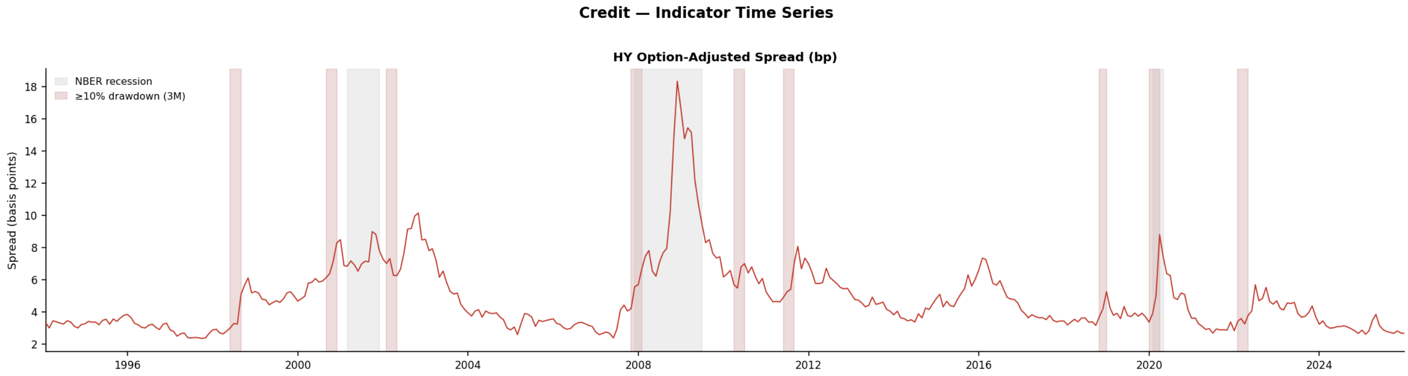

Figure 6: Credit Indicator Time Series. ICE BofA US High-Yield Option-Adjusted Spread (HY OAS), expressed in basis points. Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a $\geq$10% drawdown event). Grey shading indicates NBER recession periods.

7.1 Results

The ICE BofA US High-Yield Option-Adjusted Spread (HY OAS) is consistently the top univariate predictor in this study across short and medium horizons. It achieves AUC=0.710 at 1M (p<0.05), AUC=0.729 at 3M (p<0.05), AUC=0.663 at 6M (p<0.05), and AUC=0.617 at 12M (p<0.05), with 372 observations available from early 1994. It is selected as the Credit group representative in all four horizon-specific multi-factor models and achieves significant coefficients in both the 6M model (t=2.44, p=0.015) and 12M model (t=2.20, p=0.028).

7.2 Economic Rationale

High-yield credit spreads are a market price that directly reflects investor compensation for default risk in the riskiest segment of corporate debt. Because equity and high-yield credit markets share a common pool of investors and underlying risk, deterioration in HY credit conditions typically precedes deterioration in equity markets. The lead time is relatively short — 1–3 months — because credit markets tend to reprice corporate distress faster than equity markets, reflecting the explicit linkage of HY spreads to default probability: credit investors with capped upside react more swiftly to deteriorating fundamentals than equity holders who retain optionality on recovery. While the signal is strongest at 3M — where univariate AUC peaks — it remains statistically significant across all four horizons, suggesting HY OAS captures both the immediate repricing of corporate distress and, through its role in the multi-factor models, a persistent credit-cycle component relevant at medium and longer horizons.

The Current Market Assessment shows HY OAS as a negative driver (tight spreads) across all four horizon models as of December 2025, consistent with a credit-market environment where spreads remained compressed despite elevated equity valuations and residual inflationary pressures in services.

8 Volatility Indicators

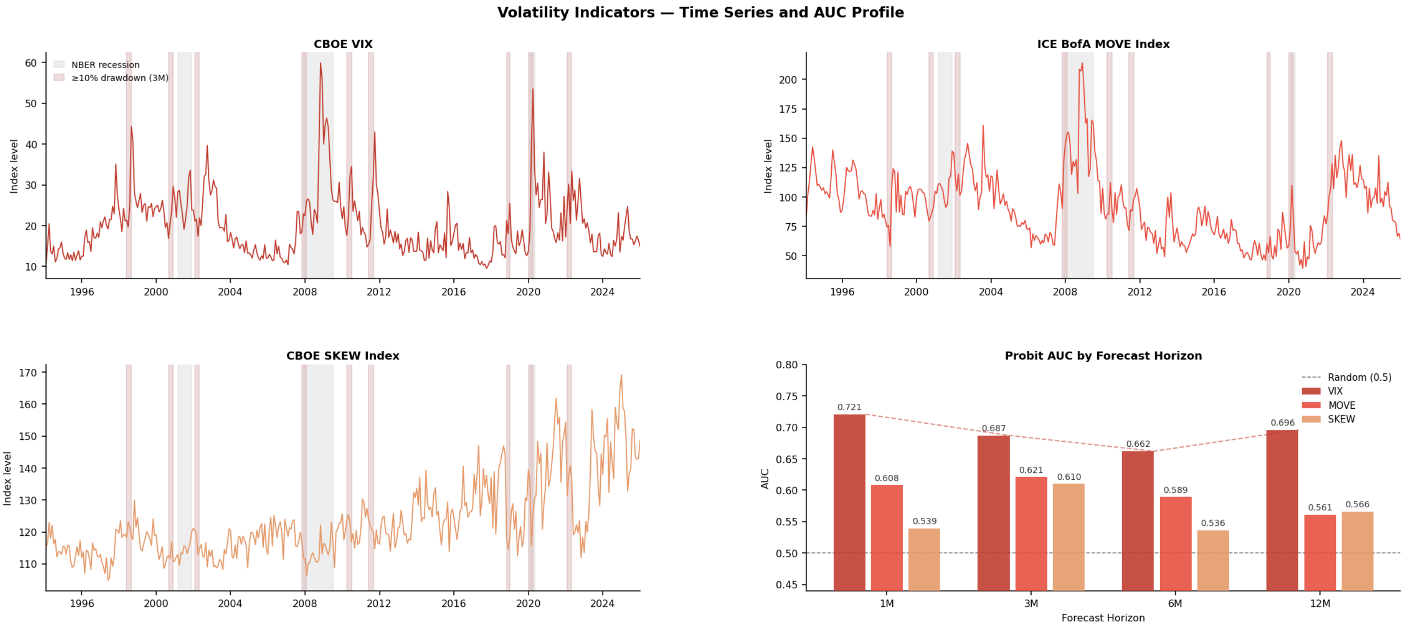

Figure 7: Volatility Indicator Time Series and AUC Profile. Top-left: CBOE VIX; top-right: ICE BofA MOVE Index; bottom-left: CBOE SKEW Index. Bottom-right: univariate probit AUC across four forecast horizons (1M, 3M, 6M, 12M) for all three volatility indicators; the dashed line traces the V-shaped AUC profile of VIX. Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a $\geq$10% drawdown event). Grey shading indicates NBER recession periods.

8.1 Results

The CBOE VIX achieves AUC=0.721 at 1M (p<0.05) and AUC=0.687 at 3M (p<0.05), making it the second-strongest short-horizon predictor overall. At 12 months, VIX regains predictive importance with AUC=0.696 (p<0.05), reflecting the distinct V-shaped pattern where elevated volatility during the initial dislocation predicts the subsequent prolonged drawdown period. The ICE BofA MOVE Index (bond market volatility) achieves AUC=0.608 at 1M and AUC=0.621 at 3M, indicating that bond market stress is a useful but secondary signal. The CBOE SKEW Index shows no significant predictive power at any horizon (AUC range: 0.536–0.610), consistent with academic evidence that implied skewness is a noisy drawdown predictor due to changing options demand dynamics.

8.2 Economic Rationale

VIX is often described as the "fear index," but its role in this study is more nuanced. At 1–3 months, elevated VIX is both a leading signal and a coincident indicator of stress: periods of genuine financial distress are typically accompanied by high volatility, and the drawdown (as measured peak-to-trough) has often already begun within the 1–3 month forward window. At 12 months, the VIX signal captures a different regime: periods of very low volatility (complacency) that precede major corrections. This structural shift means VIX exhibits two distinct predictive regimes, contributing to its relatively high AUC at both the short and long horizon.

In the multi-factor context, VIX is selected as the Volatility representative at all four horizons but shows relatively small and often insignificant individual coefficients, reflecting that its information is partially captured by HY OAS and the other indicators in the model.

9 Financial Conditions Indicators

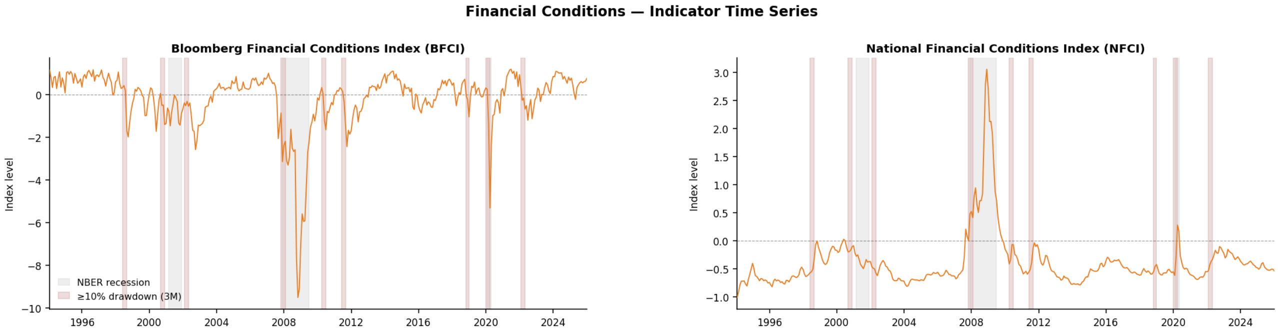

Figure 8: Financial Conditions Indicator Time Series. Bloomberg Financial Conditions Index (BFCI, left) and Chicago Fed National Financial Conditions Index (NFCI, right). For both indices, positive values indicate tighter-than-average financial conditions and negative values indicate looser-than-average conditions. Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a ≥10% drawdown event). Grey shading indicates NBER recession periods.

9.1 Bloomberg Financial Conditions Index (BFCI)

The Bloomberg Financial Conditions Index (BFCI) achieves very strong univariate AUCs: 0.723 at 1M and 0.701 at 3M, among the highest in the study. However, BFCI is excluded from all four multi-factor models due to a fundamental simultaneity concern. BFCI is constructed as a composite that incorporates equity market prices, credit spreads, and interest rates — including the very equity index we are forecasting. This creates partial circularity: when equity markets begin to decline, BFCI tightens mechanically, so its "predictive power" at short horizons partly reflects autocorrelation in equity market dynamics rather than genuine advance economic forecasting. Using BFCI in a predictive model would constitute look-ahead contamination at the 1-month horizon and partial contamination at 3 months.

This distinction matters practically: a portfolio manager cannot exploit a signal that reflects the very outcome being predicted. For this reason, we retain BFCI in the univariate analysis section as an illustrative comparison, but remove it from the multi-factor models, replacing it with the Chicago Fed National Financial Conditions Index (NFCI).

9.2 National Financial Conditions Index (NFCI)

The NFCI is published weekly by the Chicago Federal Reserve and measures credit, risk, and leverage conditions across money markets, debt and equity markets, and the traditional and shadow banking systems. Critically, it does not directly incorporate equity index levels in its construction, making it a more appropriate forecasting variable. NFCI achieves AUC=0.725 at 1M, AUC=0.673 at 3M, AUC=0.719 at 6M, and AUC=0.689 at 12M — a remarkably sustained predictive profile across all horizons. It is selected as the Financial Conditions group representative in all four models.

9.3 Economic Rationale

Financial conditions indices aggregate the tightness or ease of credit and funding markets. When financial conditions tighten — whether driven by rising credit spreads, falling asset prices, or reduced liquidity — the transmission mechanism to the real economy and corporate earnings operates over a 3–6 month horizon, explaining the strong AUC performance at those horizons. The sustained predictive power of NFCI at 12 months reflects the persistence of financial tightening episodes: once financial conditions deteriorate significantly, the recovery is gradual, and the risk of a second or deeper drawdown remains elevated for an extended period.

10 Money and Liquidity

10.1 M2 Money Supply: Raw vs QE-Adjusted

We test four money and liquidity indicators: M2 Level, M2 YoY Growth rate, a QE-adjusted M2 measure, and Federal Reserve Total Assets. M2 YoY Growth is consistently the strongest in this group, achieving AUC=0.630 at 1M, AUC=0.684 at 3M, AUC=0.667 at 6M (p<0.05), and AUC=0.695 at 12M (p<0.05). It is selected as the Money group representative in all four horizon-specific models.

The QE-adjusted M2 measure — computed as the residual from an expanding-window OLS regression of M2 YoY on Fed Assets YoY — consistently underperforms raw M2 YoY across all horizons (AUC difference: −0.043 at 12M to −0.092 at 1M). This leads us to retain raw M2 YoY Growth as the model input.

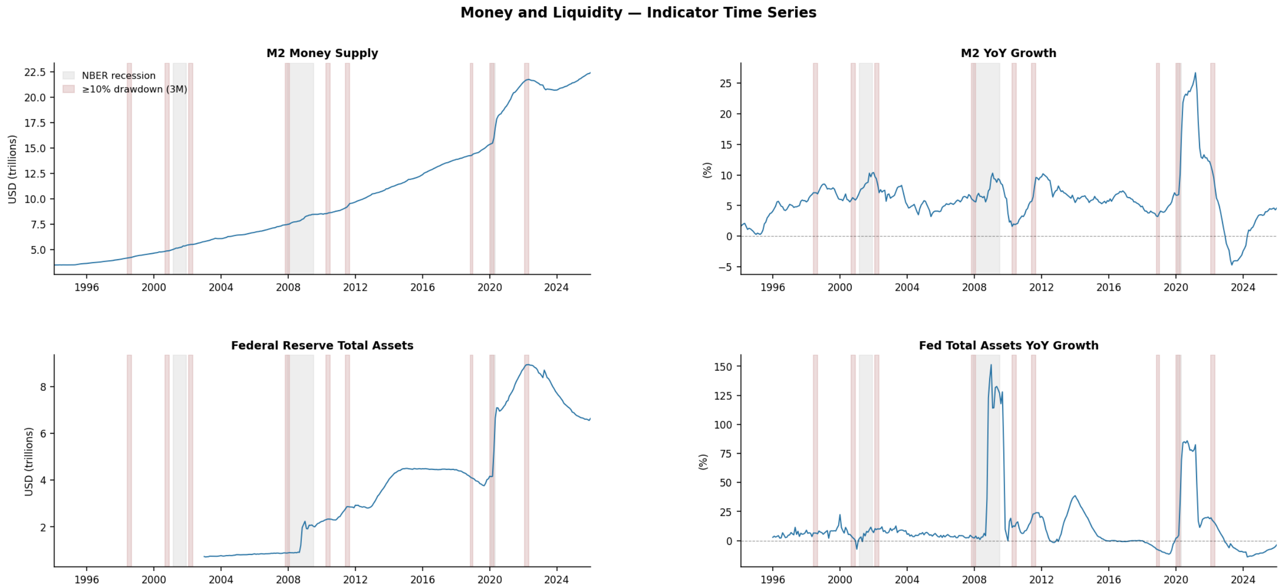

Figure 9: Money and Liquidity Indicator Time Series. Top row: M2 money supply level (USD trillions) and M2 year-on-year growth rate (%). Bottom row: Federal Reserve total assets (USD trillions) and Federal Reserve total assets year-on-year growth rate (%). Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a ≥10% drawdown event). Grey shading indicates NBER recession periods.

10.2 Dual Interpretation of M2 Growth

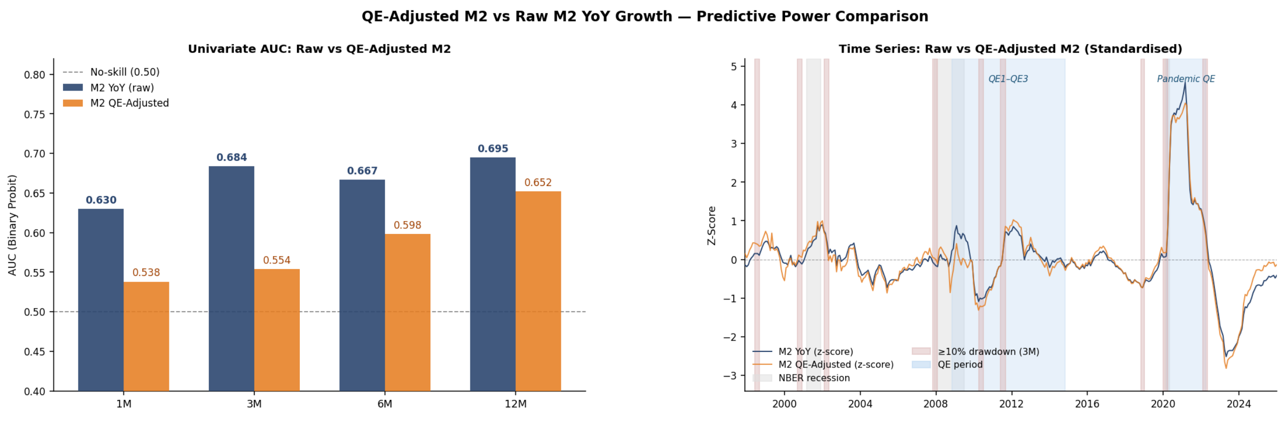

Figure 10: QE-Adjusted M2 vs Raw M2 YoY Growth — Predictive Power Comparison. Left panel: univariate probit AUC by forecast horizon for raw M2 YoY Growth and QE-Adjusted M2 (expanding-window OLS residual of M2 YoY on Fed Assets YoY); dashed line indicates the no-skill baseline (AUC = 0.50). Right panel: standardised (z-score) time series overlay of both series. Blue shading denotes Federal Reserve quantitative easing periods (QE1–QE3: November 2008–October 2014; Pandemic QE: March 2020–March 2022). Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a ≥10% drawdown event). Grey shading indicates NBER recession periods.

The consistently stronger performance of raw M2 YoY relative to the QE-adjusted series can be interpreted in two ways, with different implications for how we read the results.

Interpretation A (Mechanism): QE is part of the signal. QE and balance-sheet expansion may be a core transmission channel to asset prices. When the Federal Reserve expands its balance sheet, it injects liquidity, compresses yields, and supports equity valuations. Raw M2 YoY captures this policy-driven liquidity impulse precisely because it includes the QE component. Stripping out QE therefore removes an economically meaningful part of the relationship between monetary conditions and drawdown risk, which helps explain why the QE-adjusted series performs worse. Under this interpretation, raw M2 is the appropriate predictor.

Interpretation B (Policy reaction): QE is partly a response, not a forecast. Alternatively, QE-driven money growth may reflect the Federal Reserve's reaction function: balance-sheet expansion tends to occur when financial conditions have already deteriorated. In that case, high M2 growth may coincide with periods of stress rather than providing advance warning of future drawdowns. The raw M2 series could therefore embed a near-term crisis-response signal rather than a clean leading indicator. The QE-adjusted series, by removing this component, would in principle be a purer measure of underlying monetary growth—but in our sample it appears to lose predictive power, potentially due to limited statistical power and the difficulty of isolating a clean QE counterfactual.

We present both interpretations to avoid overstating the economic inference. In the main specifications we retain raw M2 YoY for forecasting performance, while treating the policy-reaction interpretation as an important caveat when interpreting causality.

10.3 Economic Rationale

M2 growth acquires its strongest predictive power at 12 months (AUC=0.695), reflecting the well-documented "long and variable lags" (Friedman & Schwartz, 1963) of monetary policy on the real economy. Periods of sharply decelerating M2 growth typically reflect Federal Reserve tightening which, after a transmission lag of 6–18 months, results in slower credit growth, weakening corporate revenues, and ultimately higher equity drawdown risk.

11 Consumer and Labour Market

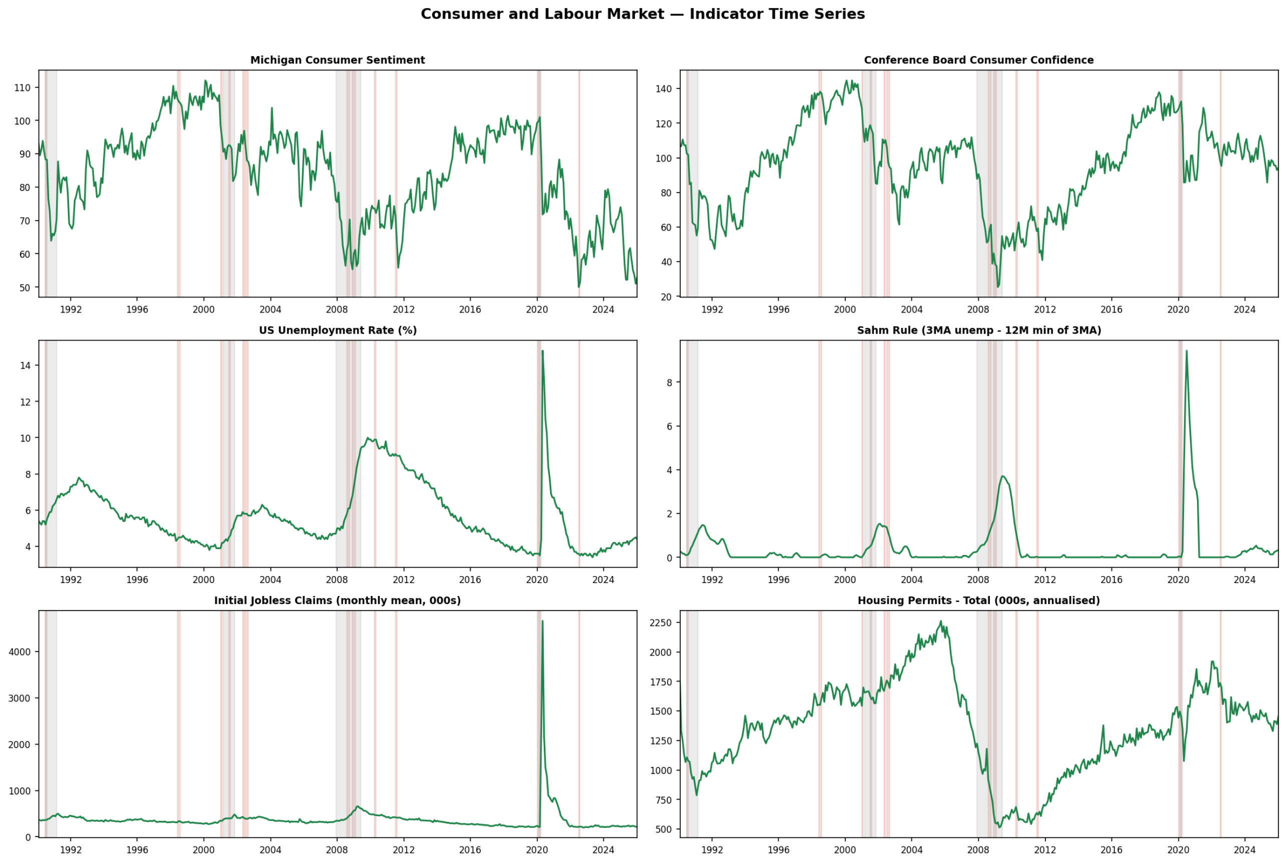

Figure 11: Consumer and Labour Market Indicator Time Series. Top row: University of Michigan Consumer Sentiment Index and Conference Board Consumer Confidence Index. Middle row: US Unemployment Rate (%) and Sahm Rule (current 3-month moving average of the unemployment rate minus the 12-month minimum of that 3-month average). Bottom row: Initial Jobless Claims (monthly mean, thousands) and US Housing Permits — Total (thousands, annualised). Red shading indicates periods where the 3-month peak-to-trough drawdown indicator equals one (i.e., a ≥10% drawdown event). Grey shading indicates NBER recession periods.

11.1 Consumer Confidence Surveys

The University of Michigan Consumer Sentiment Index and Conference Board Consumer Confidence Index both exhibit limited short-horizon predictive power (AUC ≈0.50 at 1M and 3M), consistent with survey-based indicators lagging market prices. The two surveys diverge markedly at longer horizons. Michigan Consumer Sentiment shows no meaningful predictive power at any horizon — AUC remains flat between 0.518 and 0.538 across all four horizons with no statistically significant results — reflecting its greater sensitivity to near-term consumer mood rather than underlying economic cycle turning points.

Conference Board Consumer Confidence, by contrast, shows improving predictive power at medium and longer horizons: AUC=0.587 ($p<0.10$) at 6M, rising to AUC=0.634 ($p<0.05$) at 12M. This progression reflects the slow-moving deterioration of consumer spending confidence ahead of major economic slowdowns, and the Conference Board measure's greater emphasis on labour market expectations — a channel more directly linked to equity drawdown risk than near-term consumer mood. It is selected as the Consumer & Labour group representative in the 12-month multi-factor model, where it achieves a highly significant coefficient (t=4.33, $p<0.001$).

11.2 Labour Market Indicators: Unemployment Rate vs Sahm Rule

The Sahm Rule — defined as the difference between the current 3-month moving average of the unemployment rate and its minimum over the prior 12 months — achieves AUC=0.595 at 1M, AUC=0.676 at 3M, and AUC=0.595 at 6M, compared to AUC=0.515 for the raw unemployment rate at 3M. The Sahm Rule captures the inflection point where unemployment has risen meaningfully above its recent trough, and this rate-of-change transformation is economically intuitive at short horizons: equity markets tend to respond to the acceleration of labour market deterioration rather than its absolute level.

However, this superiority is a 3-month phenomenon and should be interpreted cautiously. None of the Sahm Rule's univariate AUCs achieve statistical significance at conventional levels, and its multi-factor model coefficients are also insignificant at both horizons where it is selected — t=−0.57 (p=0.569) at 1M and t=−1.08 (p=0.279) at 3M. At the 6-month horizon, the raw unemployment rate is selected as the Consumer & Labour group representative and achieves a significant coefficient (t=−2.04, p=0.041), outperforming the Sahm Rule. At 12 months, the raw unemployment rate also records a higher univariate AUC (0.624, p<0.05) than the Sahm Rule (0.585). This horizon-dependence is itself economically meaningful: the rate-of-change captures the sharp initial inflection in labour market momentum, while the level of unemployment becomes the more reliable signal once a deterioration cycle is fully established and feeds through to corporate earnings and consumer spending. Initial Jobless Claims achieves AUC=0.631 at 3M, consistent with its role as a high-frequency barometer of labour market stress. It is not selected as the group representative in any of the four horizon models: the Sahm Rule is preferred at 1M and 3M, the raw unemployment rate at 6M, and Conference Board Consumer Confidence at 12M.

11.3 Housing Permits: Theory versus Empirics

Housing Permits (Total) shows weak and statistically insignificant predictive power across all horizons: AUC=0.515 at 1M, AUC=0.524 at 3M, AUC=0.537 at 6M, and AUC=0.568 at 12M, with none reaching conventional significance thresholds. This is disappointing relative to the theoretical case for housing as a leading equity indicator, which rests on two distinct mechanisms.

First, the fee-commitment mechanism: when a developer pulls a building permit, they commit to non-refundable permit fees, regulatory costs, and often pre-arranged construction financing. Permit issuance therefore represents a genuine forward-looking commitment to economic activity by private agents. Unlike soft-data surveys that capture intentions, permits capture actual irrevocable financial decisions. Second, the cycle-leading mechanism: the housing market is one of the most interest-rate-sensitive sectors of the economy. Housing permit applications typically slow 4–6 quarters before the broader economic cycle peaks, as rising mortgage rates and reduced affordability bite first in the residential construction pipeline.

Notably, the AUC increases steadily with the forecast horizon, reaching its highest value at 12M, consistent with a 4–6 quarter lead time as the theory predicts. The data therefore provide partial empirical support for the stated mechanism, but at a magnitude too weak and statistically too uncertain to be actionable.

The gap between the theoretical case and confirmed empirical performance highlights a core limitation of equity market prediction: sophisticated institutional investors are aware of the leading properties of housing data, and this awareness is rapidly incorporated into asset prices. By the time housing permits roll over convincingly, equity markets have often already partially discounted the slowdown. The efficient markets hypothesis, which predicts that widely-known leading indicators should be priced away, appears to operate with particular force in this sector.

This tension between the strong theoretical rationale for housing as a leading indicator and its modest empirical performance in equity return prediction is itself intellectually valuable: it reminds us that predictive power in financial markets requires either an information advantage (access to a signal before the market) or a cognitive advantage (interpreting a known signal differently). For a widely-tracked monthly series like building permits, neither advantage is readily available to investors.

11.4 Economic Rationale

The horizon profile of Consumer & Labour indicators reflects the fundamental nature of labour and consumer data: these variables measure the consequences of economic deterioration rather than its causes, and they typically lag financial market signals by one to two quarters. Results are generally weak and statistically insignificant at 1M and 3M — the Sahm Rule and Initial Jobless Claims show directional improvement at 3M but neither achieves conventional significance thresholds. The group's confirmed predictive power emerges at medium and longer horizons: the raw Unemployment Rate achieves a significant coefficient in the 6M multi-factor model (t=−2.04, p=0.041), and Conference Board Consumer Confidence achieves AUC=0.634 (p<0.05) at 12M with the largest coefficient in the long-horizon model. This progression is consistent with a regime interpretation — once labour market deterioration and consumer confidence erosion become measurable and sustained, the economy has typically entered a cycle that culminates in a major equity drawdown.

12 Cross-Indicator Comparison

Having presented univariate results for each indicator group, we now compare performance across the full set of 30 indicators using two visualisations: the AUC heatmap (Figure 12) and the R² heatmap (Figure 13), which provide a comprehensive overview of predictive patterns.

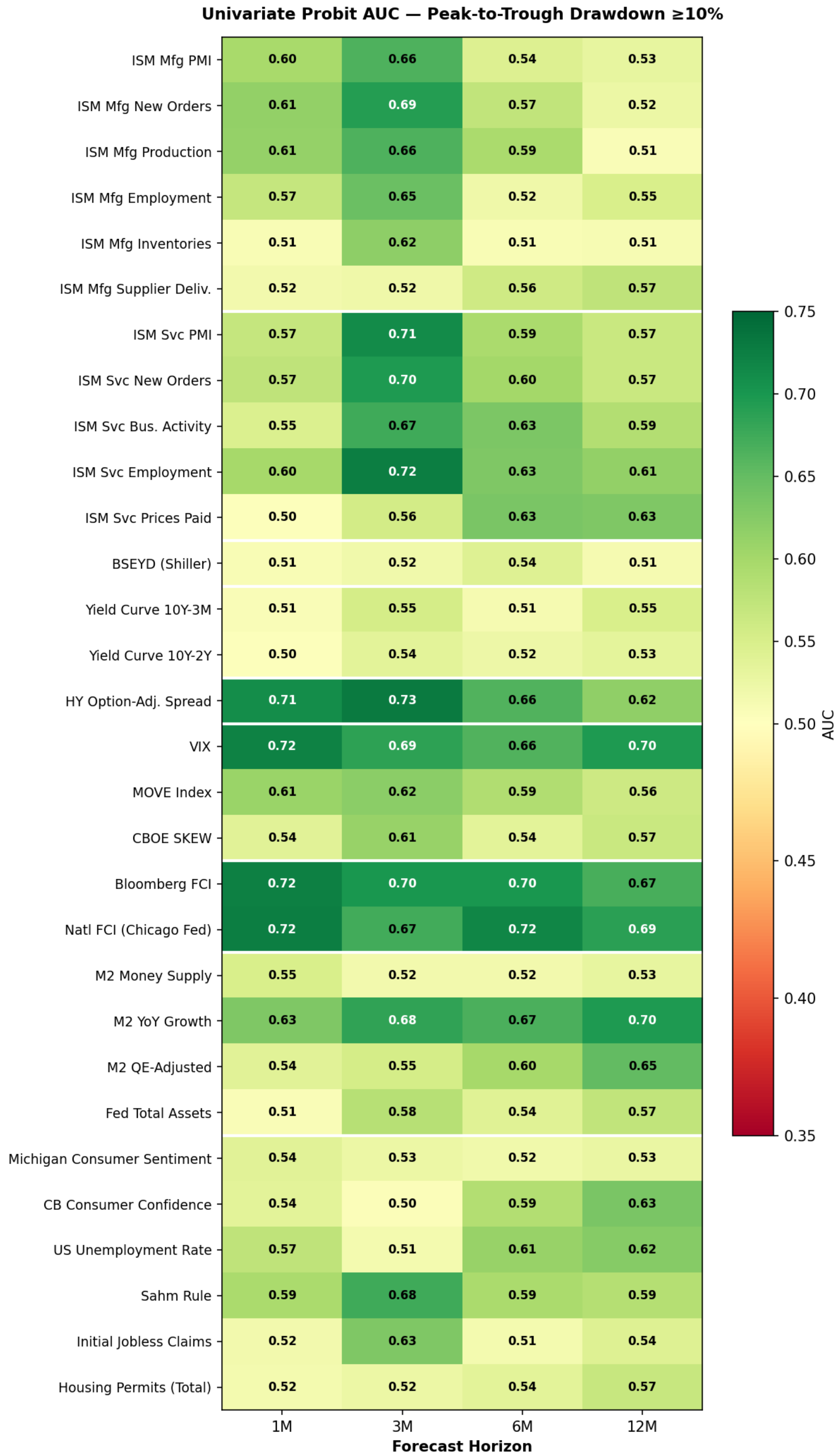

Figure 12: Univariate Probit AUC by Indicator and Horizon. Colour scale runs from red (AUC ≈0.35, weak) to green (AUC ≈0.75+, strong); values at or below 0.50 indicate no predictive power above random.

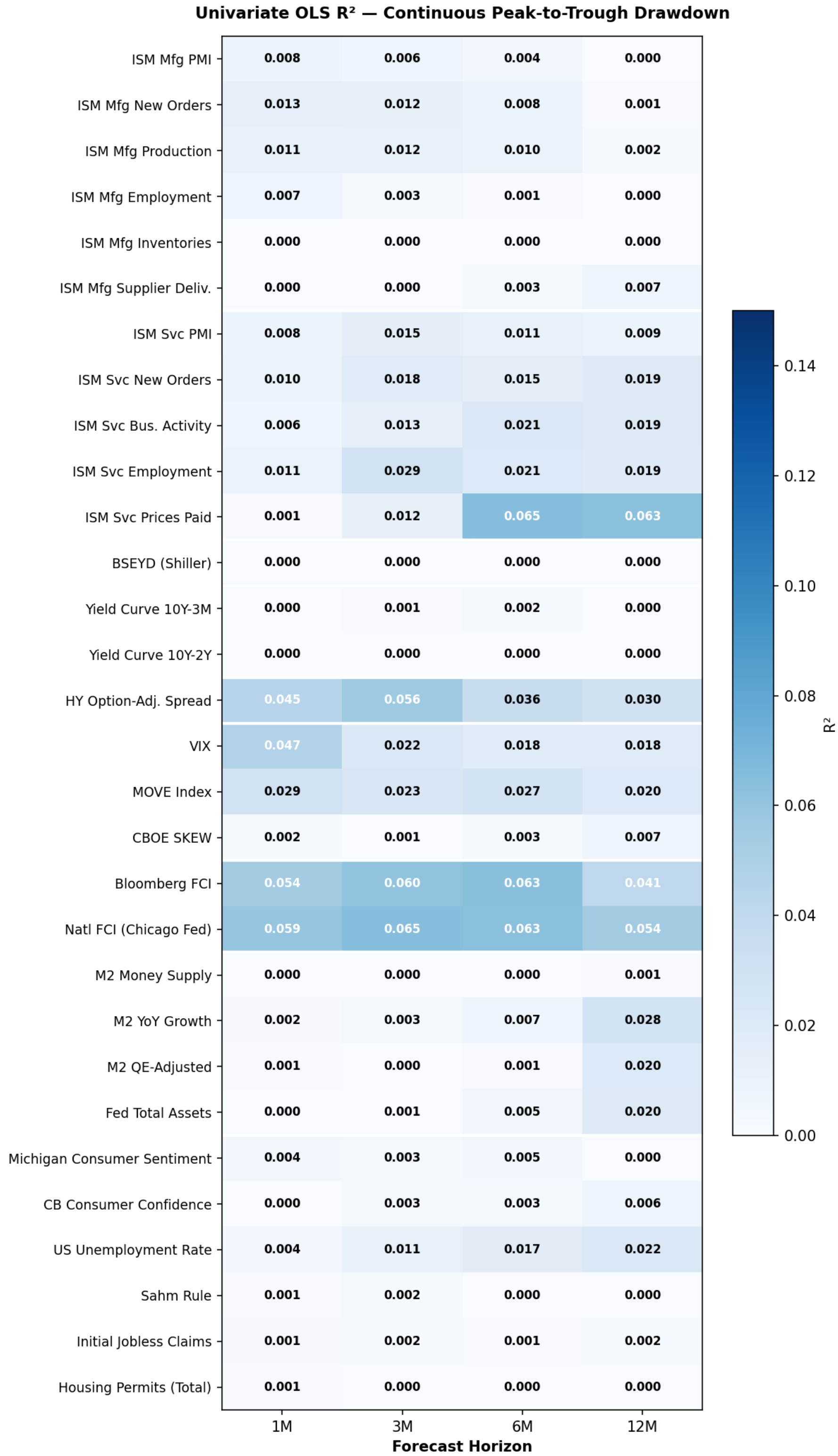

Figure 13: Univariate OLS R^2 by Indicator and Horizon. Darker blue = higher R^2.

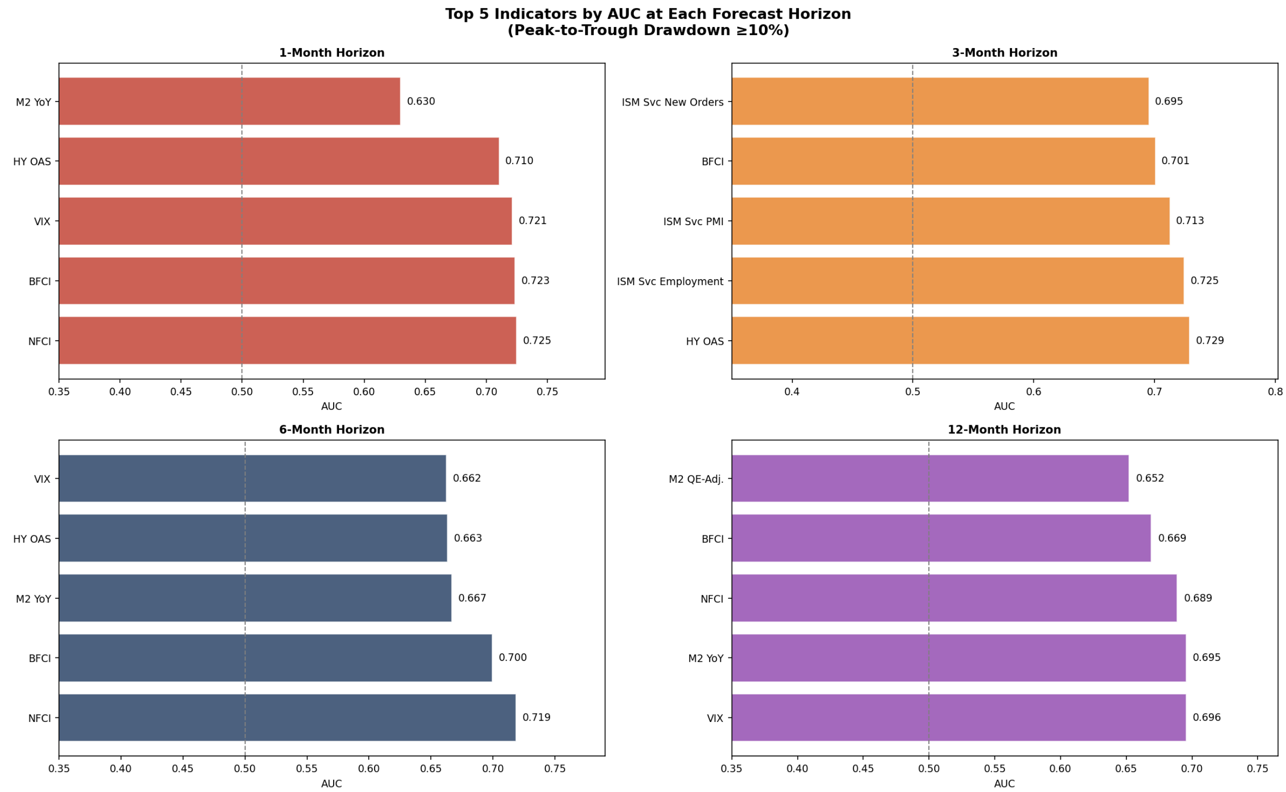

Figure 14: Top 5 Indicators by AUC at Each Forecast Horizon. Bar chart with AUC values annotated.

12.1 Key Structural Patterns

Five structural patterns emerge from the cross-indicator analysis.

First, financial market indicators dominate survey indicators at all horizons. The top univariate performers at 1M are National FCI (0.725), Bloomberg FCI (0.723), VIX (0.721), and HY OAS (0.710) — all market-price-based. Note that Bloomberg FCI is excluded from the multi-factor models due to simultaneity concerns (Section 9.1); its univariate AUC reflects its partial construction from equity market prices. Survey indicators (ISM, confidence) underperform at 1M but catch up at 3–6M horizons as real economic conditions deteriorate.

Second, the AUC profile is strongly horizon-dependent, with few indicators maintaining high predictive power across all four horizons. This motivates the four-model approach: the optimal set of predictors differs meaningfully across horizons, and imposing a single model across all horizons would sacrifice predictive efficiency.

Third, ISM Services outperforms ISM Manufacturing at 3M horizons, consistent with the greater economic weight of services in the post-2000 US economy and the faster transmission of services-sector conditions to corporate earnings in technology and financial services sectors.

Fourth, the money and liquidity group shows a distinctive long-horizon signature: M2 YoY and M2 QE-Adjusted achieve their peak predictive power at 12M, reflecting the long and variable lags of monetary transmission to equity valuations.

Fifth, the Sahm Rule records a higher univariate AUC than the raw unemployment rate at the 3M horizon (0.676 vs 0.515), suggesting that rate-of-change captures the initial inflection in labour market momentum more cleanly than the level. However, this result should be interpreted cautiously: the Sahm Rule's AUCs do not achieve statistical significance at any horizon, and at 6M and 12M the raw unemployment rate outperforms it (Section 11.2). The rate-of-change advantage is a short-horizon phenomenon rather than a general principle.

13 Modelling Challenges

13.1 Limited Drawdown Events

The main statistical challenge in this study is the scarcity of drawdown events, particularly at the 3-month horizon, where only 16 drawdown months are used in model estimation (a 4.8% base rate). With so few positive events, standard probit/logit estimates can become unstable: coefficients may be highly sensitive to small sample changes, individual predictors can appear spuriously strong, and in-sample AUCs may be overly optimistic (Peduzzi et al., 1996). We mitigate these risks in three practical ways. First, we use ridge-penalised logistic regression in the multi-factor stage ($\lambda = 1.0$) to shrink coefficients and reduce overfitting. Second, we impose a strict model size limit by including at most nine predictors (one representative per indicator group), which helps control redundancy and multicollinearity. Third, we emphasise expanding-window out-of-sample evaluation—re-estimating models through time and assessing performance on genuine forecasts—rather than relying on in-sample fit statistics.

13.2 Predictor Collinearity

Many of the indicators in this study are correlated, particularly within the financial conditions and credit groups. HY OAS, NFCI, BFCI, and VIX are all measures of financial stress and tend to spike together during market dislocations. This collinearity means that individual multi-factor model coefficients should not be interpreted as causal structural parameters, but rather as regularised weighting functions that produce a useful composite probability estimate. The ridge penalty specifically reduces the impact of collinearity by preventing any single correlated predictor from dominating the others.

13.3 Simultaneity and the BFCI Exclusion

As discussed in Section 9, the BFCI is excluded from multi-factor models due to the simultaneity concern arising from its construction incorporating equity market prices. This illustrates a broader challenge: financial market-based indicators are often the strongest univariate predictors but also the most likely to suffer from look-ahead contamination in prediction exercises. The selection of NFCI over BFCI represents a principled trade-off between predictive power and economic interpretability.

13.4 Structural Breaks and Regime Changes

The 35-year sample spans several distinct monetary policy regimes, so coefficients estimated on the full period effectively average across environments that differ materially. Prior research documents meaningful shifts in U.S. policy conduct over time, including the transition from the Volcker disinflation to the later era of conventional policy. A major break occurred after 2008, when the zero lower bound and unconventional measures (notably large-scale asset purchases) altered the usual transmission of rates, credit, and financial conditions (Bernanke & Reinhart, 2004). The forecasting literature also finds that macro-financial relationships are often structurally unstable, implying that parameters can change across regimes and weaken out-of-sample performance. Applied here, a model fit on the full 1990–2024 sample may therefore underperform within any single sub-period, especially the post-2008 unconventional policy era, which is qualitatively different and relatively short.

13.5 Expanding Window Bias

The expanding-window OOS evaluation begins with a 120-month training window in January 2000, meaning the first OOS predictions are based on data from 1990–1999 only. Indicators with short history (ISM Services: from July 1997) have fewer training observations in early OOS windows, potentially producing unstable estimates for models that include services-sector predictors. We address this by ensuring each horizon model has sufficient positive events in the training window before generating OOS predictions.

14 Multi-Factor Model Specification

We build four separate ridge-penalised logistic regression models, one per forecast horizon. At each horizon h, the model is:

$$P(\text{Drawdown}_{t,h}) = \Lambda\Big(\alpha^h + \beta_1^h\, x_1(t) + \beta_2^h\, x_2(t) + \cdots + \beta_9^h\, x_9(t)\Big)$$

where $\Lambda(\cdot)$ is the logistic function, $h \in \{1, 3, 6, 12\}$ months, and each predictor $x_k(t)$ is the z-score standardised value of the selected group representative at month t. Ridge penalisation ($\lambda = 1.0$) is applied to all slope coefficients (not the intercept).

14.1 Predictor Selection: One Representative per Group per Horizon

For each of the nine indicator groups, we select the group representative that achieves the highest univariate AUC at that specific horizon. This ensures that each model is optimised for its own horizon rather than inheriting selections made at a different horizon. BFCI is excluded from the Financial Conditions pool. The selected representatives are:

Horizon	**Selected Predictors**
1-Month	ISM Mfg New Orders, ISM Svc Employment, Yield Curve 10Y–3M, HY OAS, VIX, National FCI, BSEYD, M2 YoY, Sahm Rule
3-Month	ISM Mfg New Orders, ISM Svc Employment, Yield Curve 10Y–3M, HY OAS, VIX, National FCI, BSEYD, M2 YoY, Sahm Rule
6-Month	ISM Mfg Production, ISM Svc Prices Paid, Yield Curve 10Y–2Y, HY OAS, VIX, National FCI, BSEYD, M2 YoY, US Unemployment Rate
12-Month	ISM Mfg Supplier Deliveries, ISM Svc Prices Paid, Yield Curve 10Y–3M, HY OAS, VIX, National FCI, BSEYD, M2 YoY, CB Consumer Confidence

Table 2: Horizon-specific predictor selection. Each representative is chosen as the highest univariate AUC within its indicator group at that horizon. BFCI is excluded from the Financial Conditions pool due to simultaneity; National FCI is selected in its place across all horizons.

The shift from ISM Services Employment (selected at 1M and 3M) to ISM Services Prices Paid (selected at 6M and 12M) captures the inflation transmission channel described in Section 4. The appearance of Conference Board Consumer Confidence in the 12M model replaces the Sahm Rule, consistent with consumer sentiment being a slower moving but sustained signal at the longer horizon.

14.2 Out-of-Sample Evaluation

Out-of-sample performance is evaluated via an expanding-window walk-forward procedure. For each month t from January 2000 onward (where the training window [1990-01, $t-1$] contains at least 120 observations), we:

1. estimate the ridge logit on training data only;
2. predict the drawdown probability for month t using the estimated coefficients and the predictor values observed at t;
3. advance t by one month.

This produces approximately 300 OOS predictions per model without any forward-looking information.

15 Model Results

Table 3a presents coefficients for the 1-Month Short-Horizon Stress Detection Model. Table 3b presents the 3-Month Near-Term Leading Indicator Model. Tables 3c and 3d present the 6-Month and 12-Month models respectively. Figures 15a and 15b provide a comprehensive two-row visual summary: the top row shows ROC curves (IS and OOS) for each model; the bottom row shows the predicted drawdown probability time series with OOS expanding-window predictions overlaid in a contrasting colour.

15.1 Short-Horizon Stress Detection Model (h=1M)

The 1-Month model achieves in-sample AUC=0.736 and OOS AUC=0.600 (McFadden R^2=0.107), evaluated on approximately 300 OOS months from January 2000. The threshold definition for h=1 is a 5% single-month decline (anchor-at-t). BSEYD is the only statistically significant predictor (t=-2.19, p=0.028), with a negative coefficient indicating that expensive equity valuations increase drawdown probability. NFCI and HY OAS have positive (tightening → higher drawdown risk) coefficients consistent with expectations, but are not individually significant in the presence of the other predictors, reflecting collinearity.

Predictor	**Coef. (β)**	**Std. Err.**	**t-stat**	**p-value**	**Direction**
ISM Mfg New Orders	−0.216	0.302	−0.72	0.475	Lower orders → higher drawdown prob.
ISM Svc Employment	0.352	0.327	1.08	0.281	Higher employment → higher drawdown prob.
Yield Curve 10Y–3M	0.256	0.251	1.02	0.308	Steeper curve → higher drawdown prob.
HY OAS	0.517	0.371	1.39	0.163	Wider spread → higher drawdown prob.
VIX	0.271	0.234	1.16	0.246	Higher VIX → higher drawdown prob.
National FCI	0.515	0.484	1.06	0.288	Tighter FCI → higher drawdown prob.
BSEYD (Shiller)	−0.499	0.228	−2.19	0.028**	Expensive equity → higher drawdown prob.
M2 YoY Growth	0.005	0.225	0.02	0.982	(Negligible at 1M)
Sahm Rule	−0.176	0.308	−0.57	0.569	(Absorbed by FCI at 1M)

Ridge-penalised logistic regression ($\lambda = 1.0$). h = 1M drawdown = single-month decline ≥ 5% (anchor-at-t).
N = 330 obs.; Drawdown months = 38; Base rate = 11.5%. IS AUC = 0.736; OOS AUC = 0.600; McFadden R^2 = 0.107.
** $p < 0.05$; * $p < 0.10$.

Table 3a: Short-Horizon Stress Detection Model (h = 1M, Drawdown ≥ 5% Monthly Decline).

15.2 Near-Term Leading Indicator Model (h=3M)

The 3-Month model achieves IS AUC=0.785 and OOS AUC=0.694, with McFadden R^2=0.147. Only 16 drawdown months at this horizon (base rate 4.8% under the peak-to-trough 10% threshold) limits individual coefficient significance. No predictor achieves p<0.05 significance individually, though the collective model discriminates well. The negative signs on ISM Mfg New Orders and ISM Svc Employment indicate that contractionary survey readings are associated with higher near-term drawdown probability, consistent with economic logic.

Predictor	**Coef. (β)**	**Std. Err.**	**t-stat**	**p-value**	**Direction**
ISM Mfg New Orders	−0.451	0.401	−1.13	0.260	Contractionary survey → higher drawdown prob.
ISM Svc Employment	−0.595	0.439	−1.36	0.175	Employment weakness → higher drawdown prob.
Yield Curve 10Y–3M	0.387	0.368	1.05	0.293	Steeper curve → higher drawdown prob.
HY OAS	0.531	0.484	1.10	0.273	Wider spread → higher drawdown prob.
VIX	−0.332	0.379	−0.88	0.380	VIX mean-reverts vs 3M (mixed)
National FCI	−0.064	0.588	−0.11	0.913	(Absorbed by other predictors)
BSEYD (Shiller)	−0.319	0.327	−0.98	0.329	Expensive equity risk
M2 YoY Growth	−0.028	0.349	−0.08	0.937	(Negligible at 3M)
Sahm Rule	−0.470	0.433	−1.08	0.279	Falling unemployment → higher drawdown prob.

Ridge-penalised logistic regression ($\lambda = 1.0$). h = 3M peak-to-trough drawdown ≥ 10%.
N = 330 obs.; Drawdown months = 16; Base rate = 4.8%. IS AUC = 0.785; OOS AUC = 0.694; McFadden R^2 = 0.147.
Note: individual coefficients lack significance due to very few drawdown events (16); the collective model discriminates well.
** $p < 0.05$; * $p < 0.10$.

Table 3b: Near-Term Leading Indicator Model (h = 3M, Peak-to-Trough ≥ 10%).

15.3 Medium-Term Leading Indicator Model (h=6M)

The 6-Month model is the strongest performer overall: IS AUC=0.778 and OOS AUC=0.737, with McFadden R^2=0.183. Five predictors achieve statistical significance at the 5% level: ISM Svc Prices Paid (t=5.37, p<0.001), Yield Curve 10Y-2Y (t=2.39, p=0.017), HY OAS (t=2.44, p=0.015), BSEYD (t=-2.93, p=0.003), and US Unemployment Rate (t=-2.04, p=0.041). The ISM Services Prices Paid coefficient (0.965) is the largest in the study, reflecting that elevated services price pressures at the 6-month horizon capture the full spectrum of the monetary tightening transmission channel.

Predictor	**Coef. (β)**	**Std. Err.**	**t-stat**	**p-value**	**Direction**
ISM Mfg Production	−0.451	0.244	−1.85	0.064*	Manufacturing contraction risk
ISM Svc Prices Paid	0.965	0.180	5.37	0.000**	Inflation pressure → higher drawdown prob.
Yield Curve 10Y–2Y	0.607	0.254	2.39	0.017**	Steeper curve → higher drawdown prob.
HY OAS	0.888	0.364	2.44	0.015**	Wider spread → higher drawdown prob.
VIX	−0.118	0.253	−0.47	0.640	(Absorbed at 6M)
National FCI	−0.041	0.448	−0.09	0.926	(Absorbed by credit/prices)
BSEYD (Shiller)	−0.621	0.212	−2.93	0.003**	Expensive equity → higher drawdown prob.
M2 YoY Growth	0.057	0.156	0.37	0.714	(Indirect at 6M)
US Unemployment Rate	−0.487	0.238	−2.04	0.041**	Falling unemployment → higher drawdown prob.

Ridge-penalised logistic regression ($\lambda = 1.0$). h = 6M peak-to-trough drawdown $\geq$ 10%.
N = 330 obs.; Drawdown months = 61; Base rate = 18.5%. IS AUC = 0.778; OOS AUC = 0.737; McFadden R^2 = 0.183.
** $p < 0.05$; * $p < 0.10$.

Table 3c: Medium-Term Leading Indicator Model (h = 6M, Peak-to-Trough $\geq$ 10%).

15.4 Long-Term Leading Indicator Model (h=12M)

The 12-Month model achieves IS AUC=0.742 and OOS AUC=0.684, with McFadden R^2=0.151. Four predictors are highly significant: ISM Svc Prices Paid (t=5.09, p<0.001), Yield Curve 10Y–3M (t=2.04, p=0.042), HY OAS (t=2.20, p=0.028), and Conference Board Consumer Confidence (t=4.33, p<0.001). The Conference Board coefficient of 0.901 is among the largest in the study: note that this indicator is entered standardised, so a one standard deviation increase in consumer confidence is associated with a large increase in 12-month drawdown probability. This counter-intuitive direction reflects the late-cycle complacency effect: peak consumer confidence often occurs near market peaks, making it a contrarian signal at long horizons.

Predictor	**Coef. (β)**	**Std. Err.**	**t-stat**	**p-value**	**Direction**
ISM Mfg Supplier Deliv.	−0.604	0.373	−1.62	0.105	Easing bottlenecks → late-cycle risk
ISM Svc Prices Paid	0.904	0.178	5.09	0.000**	Inflation pressure → higher drawdown prob.
Yield Curve 10Y–3M	0.429	0.211	2.04	0.042**	Steeper curve → higher drawdown prob.
HY OAS	0.671	0.306	2.20	0.028**	Wider spread → higher drawdown prob.
VIX	−0.135	0.206	−0.66	0.512	(Late-cycle VIX often low)
National FCI	0.467	0.418	1.12	0.264	(Positive at long horizon)
BSEYD (Shiller)	−0.058	0.168	−0.35	0.729	(Absorbed at 12M)
M2 YoY Growth	0.249	0.139	1.79	0.073*	Higher M2 → higher drawdown prob.
CB Consumer Confidence	0.901	0.208	4.33	0.000**	Complacency at cycle peak

Ridge-penalised logistic regression ($\lambda = 1.0$). h = 12M peak-to-trough drawdown $\geq$ 10%.
N = 330 obs.; Drawdown months = 86; Base rate = 26.1%. IS AUC = 0.742; OOS AUC = 0.684; McFadden R^2 = 0.151.
** $p < 0.05$; * $p < 0.10$.

Table 3d: Long-Term Leading Indicator Model (h = 12M, Peak-to-Trough $\geq$ 10%).

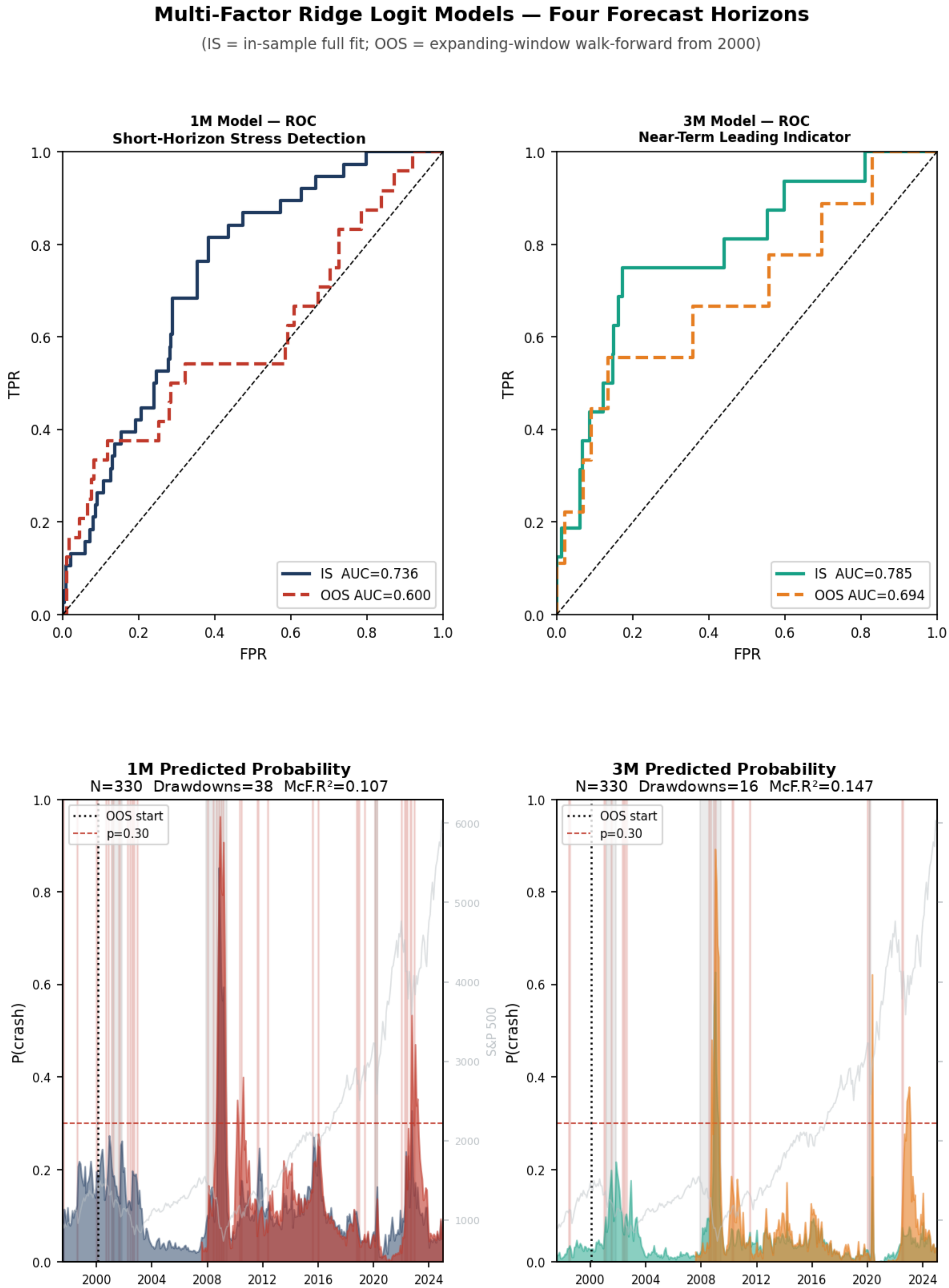

Figure 15a: Multi-Factor Ridge Logit Models — Short and Near-Term Horizons (1M and 3M). Top row: in-sample (IS) and expanding-window out-of-sample (OOS) ROC curves for the 1-Month Short-Horizon Stress Detection Model (left) and 3-Month Near-Term Leading Indicator Model (right). Bottom row: predicted drawdown probability time series. Black dotted line = OOS start (January 2000); dashed red line = $p = 0.30$ alert threshold. Grey shading denotes NBER recession periods.

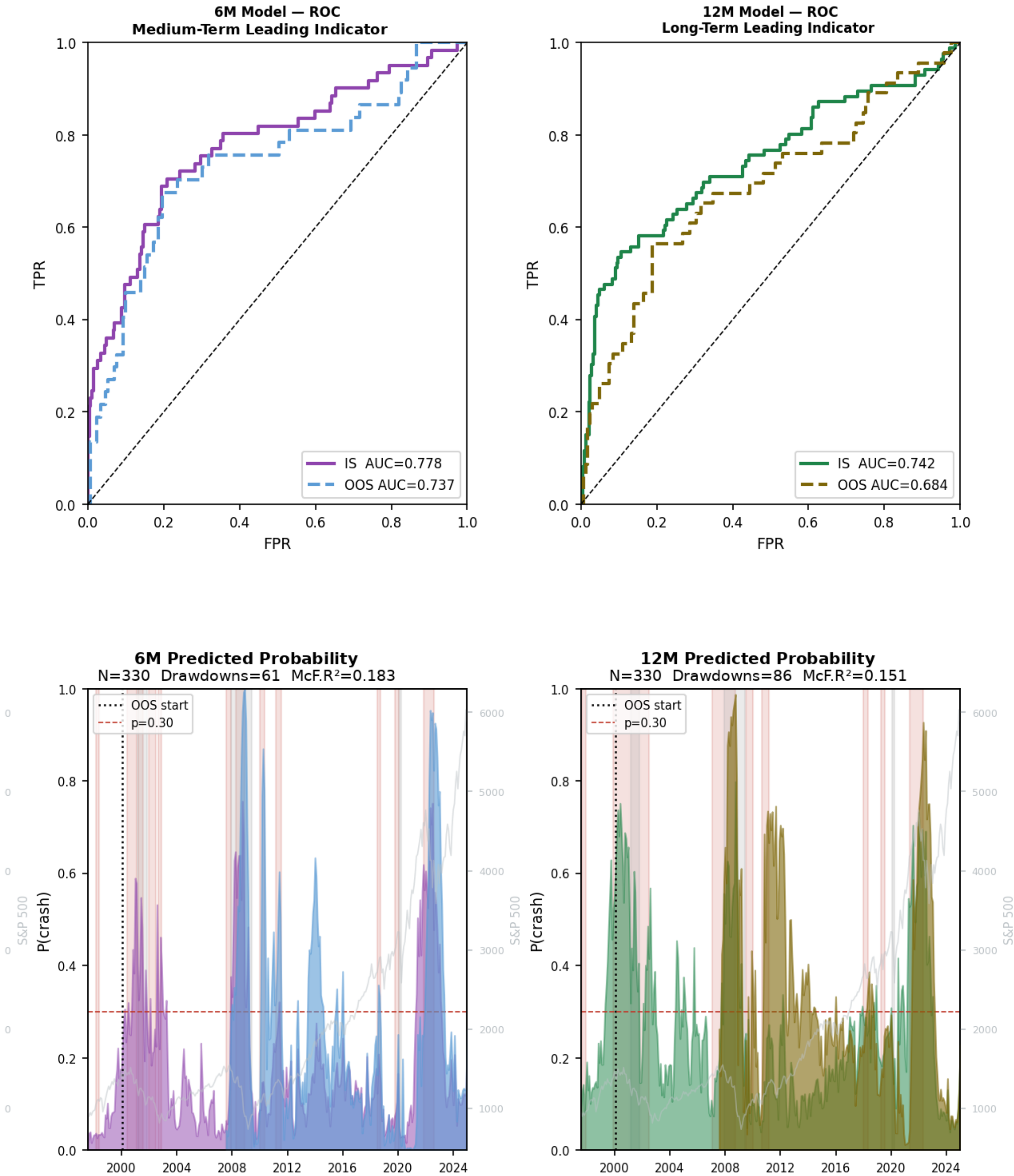

Figure 15b: Multi-Factor Ridge Logit Models — Medium and Long-Term Horizons (6M and 12M). Top row: IS and OOS ROC curves for the 6-Month Medium-Term Leading Indicator Model (left) and 12-Month Long-Term Leading Indicator Model (right). Bottom row: predicted drawdown probability time series. The 6-Month model achieves the highest OOS AUC across all four horizons (0.737), consistent with the 2–3 quarter transmission lag of macro-financial signals. All conventions as in Figure 15a.

15.5 Summary of Model Performances

Across the four ridge-logit specifications, predictive performance is strongly horizon-dependent. The 6-month model is the clear top performer, achieving OOS AUC = 0.737 and the highest pseudo-R^2, with the richest set of statistically significant predictors. This pattern is consistent with a macro-financial transmission lag of roughly two to three quarters, over which tightening financial conditions, restrictive policy dynamics, and weakening labour and demand conditions can jointly propagate into equity drawdowns. In the 6M specification, the joint significance of Services Prices Paid, the yield curve, HY OAS, unemployment, and valuation (BSEYD) supports an economically coherent narrative: drawdown risk rises when inflation pressure keeps policy restrictive, credit compensation widens, and labour-market softening becomes visible—particularly when valuations offer little cushion.

The 3-month model achieves respectable discrimination (OOS AUC = 0.694) despite extreme event scarcity at this horizon (only 16 positive months in the sample). In this setting, coefficient-level inference is naturally weak: with so few positives, individual predictors rarely reach conventional significance even when the combined model ranks risk meaningfully. Operationally, however, this horizon should be treated as a high-noise tactical signal. Under an illustrative probability threshold (e.g., $p > 0.30$), the model's precision is low relative to the other horizons, implying a high false-alarm rate that would require either (i) sustained signals over multiple months, or (ii) a higher threshold and additional confirmation filters to be usable in practice.

The 12-month model (OOS AUC = 0.684) behaves less like a short-run stress detector and more like a late-cycle regime identification tool. Its most informative predictors—Services Prices Paid, the 10Y–3M curve, HY OAS, and Conference Board confidence—map naturally into a longer-horizon mechanism in which persistent inflation pressure sustains restrictive policy, credit conditions gradually tighten, and late-cycle sentiment can act as a contrarian peak signal. The positive coefficient on consumer confidence is best interpreted cautiously: it should not be read as confidence *causing* drawdowns, but rather as capturing the empirical tendency for sentiment to peak late in the cycle, when downside vulnerability is elevated.

Finally, the 1-month model (OOS AUC = 0.600) is the weakest performer, which is expected given the near-random-walk behaviour of short-horizon equity returns and the definitional change at $h = 1$ (a $\geq 5\%$ single-month decline rather than an intra-window peak-to-trough drawdown). In this very short window, macro-financial variables are less able to "lead" the outcome; instead, the model functions primarily as a risk-temperature gauge, with valuation (BSEYD) emerging as the lone individually significant term.

Overall, the horizon comparison implies a practical hierarchy: the 6M and 12M horizons are the most useful for macro-informed risk management, offering higher discrimination and more coherent economic structure; the 3M horizon can be informative but should be treated as a noisier tactical layer; and the 1M horizon is best interpreted as short-horizon stress monitoring rather than true advance forecasting.

Table 4: Comprehensive Univariate Probit Results: AUC and McFadden R^2 by Indicator and Horizon.

Indicator	**AUC 1M**	R^2 **1M**	**AUC 3M**	R^2 **3M**	**AUC 6M**	R^2 **6M**	**AUC 12M**	R^2 **12M**	N
ISM Manufacturing									
ISM Mfg PMI	0.596**	0.022	0.664**	0.048	0.544	0.002	0.530	0.001	420
ISM Mfg New Orders	0.615**	0.033	0.692**	0.063	0.573	0.007	0.524	0.002	420
ISM Mfg Production	0.612**	0.029	0.665**	0.052	0.594**	0.011	0.507	0.000	420
ISM Mfg Employment	0.569*	0.012	0.646*	0.026	0.520	0.000	0.550	0.002	420
ISM Mfg Inventories	0.509	0.000	0.618	0.011	0.509	0.000	0.510	0.000	420
ISM Mfg Supplier Deliv.	0.517	0.001	0.520	0.004	0.561*	0.008	0.574**	0.016	420
ISM Services									
ISM Svc PMI	0.570*	0.012	0.713**	0.072	0.594	0.005	0.565	0.002	330
ISM Svc New Orders	0.575*	0.016	0.695**	0.062	0.601	0.008	0.566	0.006	330
ISM Svc Bus. Activity	0.547	0.007	0.674**	0.051	0.631*	0.010	0.587*	0.010	330
ISM Svc Employment	0.598*	0.014	0.725**	0.081	0.629**	0.018	0.614**	0.014	330
ISM Svc Prices Paid	0.502	0.000	0.556	0.008	0.633**	0.047	0.630**	0.039	330
Valuation									
BSEYD (Shiller)	0.509	0.000	0.518	0.009	0.542	0.003	0.513	0.001	420
Yield Curve									
Yield Curve 10Y–3M	0.508	0.000	0.554	0.005	0.512	0.001	0.550	0.003	420
Yield Curve 10Y–2Y	0.504	0.000	0.537	0.004	0.522	0.000	0.535	0.002	420
Credit									
HY OAS	0.710**	0.073	0.729**	0.105	0.663**	0.033	0.617**	0.016	372
Volatility									
VIX	0.721**	0.073	0.687**	0.043	0.662**	0.026	0.696**	0.025	420
MOVE Index	0.608**	0.032	0.621**	0.042	0.589**	0.018	0.561	0.005	420
CBOE SKEW	0.539	0.003	0.610	0.016	0.536	0.001	0.566	0.001	420
Financial Conditions									
Bloomberg FCI	0.723**	0.070	0.701**	0.071	0.700**	0.041	0.669**	0.018	420
National FCI	0.725**	0.069	0.673**	0.068	0.719**	0.040	0.689**	0.026	420
Money & Liquidity									
M2 Level	0.550	0.001	0.516	0.004	0.517	0.000	0.530	0.000	420
M2 YoY Growth	0.630	0.005	0.684	0.006	0.667**	0.012	0.695**	0.033	420
M2 QE-Adjusted	0.538	0.000	0.554	0.000	0.598	0.001	0.652**	0.016	314
Fed Total Assets	0.507	0.001	0.582	0.016	0.540	0.004	0.573**	0.015	361
Consumer & Labour Market									
Michigan Sentiment	0.538	0.007	0.525	0.005	0.518	0.001	0.526	0.002	420
CB Consumer Conf.	0.537	0.001	0.505	0.001	0.587*	0.008	0.634**	0.024	420
US Unemployment Rate	0.575	0.005	0.515	0.000	0.611*	0.012	0.624**	0.014	420
Sahm Rule	0.595	0.000	0.676	0.007	0.595	0.000	0.585	0.002	420
Initial Jobless Claims	0.516	0.002	0.631	0.000	0.514	0.004	0.541	0.001	420
Housing									
Housing Permits (Total)	0.515	0.004	0.524	0.006	0.537	0.000	0.568	0.002	420

Univariate probit models. AUC = area under the ROC curve. R^2 = McFadden pseudo-R^2.
Significance based on probit z-test on slope coefficient. ** $p < 0.05$; * $p < 0.10$.
N reflects data availability; ISM Services series begin July 1997 ($N = 330$). HY OAS begins January 1994 ($N = 372$).

16 Historical Drawdown Episode Analysis

To provide a practically-oriented evaluation of model performance, we analyse nine major equity market dislocations since 1990. For each episode, we assess whether each of the four horizon-specific models generated a warning signal - defined as a predicted probability exceeding the 0.30 threshold - in the months prior to the drawdown onset. Pre-2000 episodes fall within the in-sample period for the expanding-window evaluation; post-2000 episodes use genuine out-of-sample predictions.

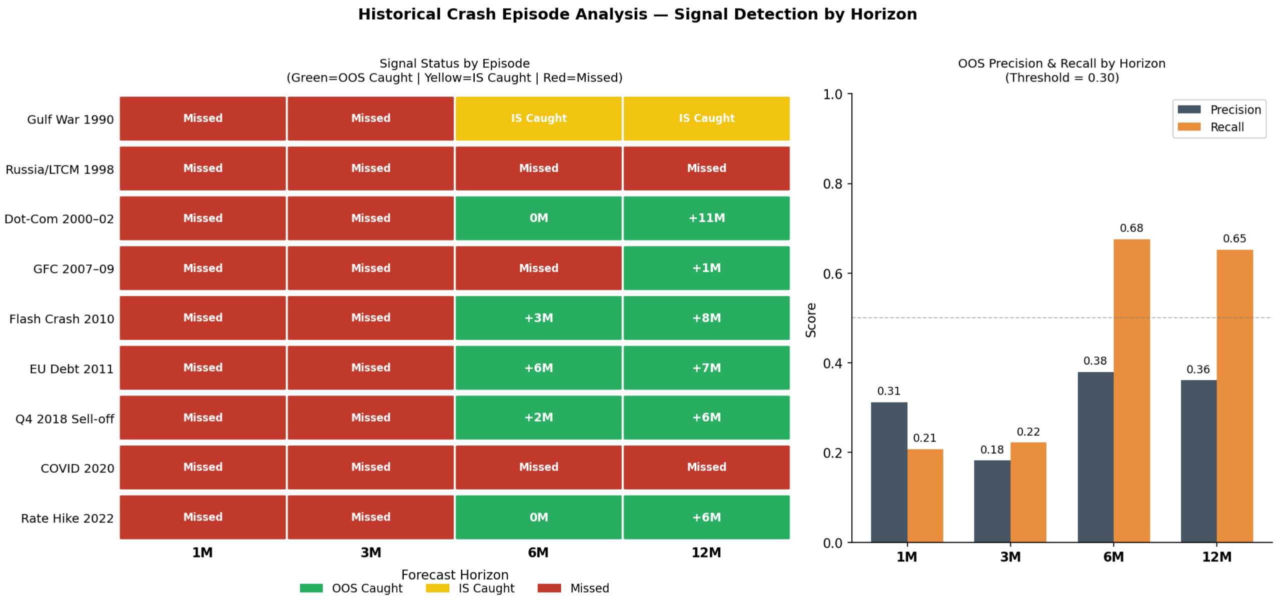

Figure 16: Historical Drawdown Episode Analysis. Left panel: signal status matrix (green = OOS caught, yellow = in-sample caught, red = missed); numbers in green cells indicate months of advance warning prior to episode onset (e.g. +6M means the model's predicted probability first crossed the 0.30 threshold six months before the drawdown began; 0M indicates the threshold was crossed in the same month as onset). Right panel: OOS precision and recall by horizon at the 0.30 probability threshold.

16.1 Episode Commentary

Gulf War Recession (1990): This episode falls in the in-sample period for all models. The 6-month and 12-month models tend to flag this period given the broad deterioration in ISM data, financial conditions, and yield curve dynamics leading into the Gulf conflict. The role of oil-price shocks in precipitating recessionary conditions is well documented (Hamilton, 2003).

Russia/LTCM Crisis (1998): An idiosyncratic financial contagion event. The abrupt onset of the Russian default and LTCM implosion meant that few macro indicators provided advance warning. Jorion (2000) documents how the speed and cross-asset nature of the contagion overwhelmed conventional risk models, illustrating a known limitation of macro-based forecasting approaches.

Dot-Com Bust (2000–2002): The prolonged nature of this drawdown (peak March 2000, trough September 2002) provides multiple sequential signal windows. The 6M and 12M models, using OOS predictions from 2000 onwards, demonstrate their ability to maintain elevated probability signals throughout the extended drawdown period. The valuation excesses preceding this episode are extensively documented in the literature (Shiller, 2000; Pastor and Veronesi, 2006).

Great Financial Crisis (2007–2009): The most severe episode in the sample. The credit-spread and financial conditions signals (HY OAS, NFCI) provided early warning through 2007. The 6M model shows strong recall for this episode, consistent with its overall high OOS AUC. Brunnermeier (2009) provides a detailed account of how deteriorating credit and liquidity conditions propagated through the financial system well before the acute phase of the crisis.

Flash Crash (2010) and EU Debt Crisis (2011): These shorter episodes are among the most challenging to predict. The Flash Crash (May 2010) was a liquidity-driven microstructure event with no clear macro antecedents; Kirilenko et al. (2017) attribute it primarily to high-frequency trading dynamics rather than any deterioration in underlying fundamentals. The EU Debt Crisis (2011) was driven by European sovereign contagion, which is only partially captured by US macro indicators (Lane, 2012).

Q4 2018 Sell-Off: Driven by Federal Reserve tightening and trade war escalation. The yield curve signal (approaching inversion) and HY OAS widening provided preliminary warning in the months prior, consistent with the predictive role of financial conditions documented by Gilchrist and Zakrajšek (2012).

COVID-19 Crash (2020): The most rapid drawdown in the sample (January to March 2020). By construction, no macro-based model can provide advance warning of a novel pandemic event. Models flagged this period only after the market had already begun to fall, illustrating the fundamental boundary of macro forecasting. Baker et al. (2020) document the unprecedented speed and scale of the market reaction relative to all prior shock events.

Rate Hike Drawdown (2022): Unlike COVID-19, this was a macro-driven correction with clear antecedents: ISM Services Prices Paid remained elevated throughout 2021, the yield curve began inverting in late 2021, and consumer confidence was showing signs of late-cycle complacency. The 12-month model, which most heavily weights these indicators, shows its best episode-level performance here. The inflationary dynamics and subsequent aggressive tightening cycle that drove this correction are consistent with the monetary transmission mechanisms described in Clarida, Galí, and Gertler (2000).

16.2 Precision and Recall

Table 5 presents OOS precision and recall for each model at the 0.30 probability threshold.

Table 5: OOS Precision and Recall by Horizon Model (Threshold = 0.30)

Horizon	Model	Threshold	TP	FP	FN	Precision	Recall
1M	Short-Horizon Stress Detection	0.30	17	37	65	0.315	0.208
3M	Near-Term Leading Indicator	0.30	2	9	7	0.182	0.222
6M	Medium-Term Leading Indicator	0.30	48	79	23	0.379	0.676
12M	Long-Term Leading Indicator	0.30	60	106	32	0.361	0.652

TP = true positives (drawdown months correctly signalled).
FP = false alarms (signal fired, no subsequent drawdown within horizon).
FN = missed events (drawdown months with no prior signal).
Precision = TP/(TP + FP); Recall = TP/(TP + FN).
OOS evaluation period: January 2000 onwards.

The 6-Month model achieves the highest recall (0.676), meaning it correctly identifies 67.6% of drawdown months in the OOS period. Precision (0.379) indicates that approximately 38% of the months flagged as high-risk were followed by a drawdown within 6 months — or, equivalently, 62% were false alarms. The 1-Month model shows lower recall (0.208) due to the inherent difficulty of very short-horizon prediction and the definitional difference in the drawdown threshold. These precision and recall statistics should be interpreted in context: a portfolio manager does not need to predict every drawdown month independently. If the model flags a sustained elevated-risk regime (several consecutive months above threshold), the signal should be interpreted as a warning window rather than a single-month prediction.

17 Current Market Assessment

Applying the four horizon-specific models to the most recent available predictor data (December 2025) yields the probability estimates shown in Table 6. Full-sample fitted coefficients are used throughout; confidence intervals are constructed via the delta method. These estimates are for analytical purposes only and do not constitute investment advice.

Table 6: Current Market Assessment (as of December 2025)

Horizon	**Model**	$\hat{P}$	**CI Lower**	**CI Upper**	**Key Drivers**
1M	Short-Horizon Stress Detection	5.0%	0.9%	9.0%	HY OAS (tight) BSEYD (elevated) Natl FCI (accommodative)
3M	Near-Term Leading Indicator	2.0%	0.0%	4.4%	HY OAS (tight) Yield Curve (flat) BSEYD (elevated)
6M	Medium-Term Leading Indicator	12.4%	5.2%	19.6%	HY OAS (tight) BSEYD (elevated) Unemployment Rate (rising)
12M	Long-Term Leading Indicator	8.2%	2.7%	13.7%	HY OAS (tight) Yield Curve (flat) ISM Svc Prices Paid (elevated)

95% confidence intervals via delta method. Alert threshold: $\hat{P} > 0.30$.
Estimates based on December 2025 predictor values; not investment advice.

All four horizons sit well below the 0.30 alert threshold. The 1M and 3M models return probabilities of 5.0% and 2.0% respectively, indicating low near-term stress. The 6M model produces the highest estimate at 12.4% (95% CI: 5.2%–19.6%), with the 12M model at 8.2% (2.7%–13.7%). Relative to the December 2024 assessment, probabilities have declined materially at all horizons, reflecting improved credit and labour-market conditions over the course of 2025.

Across all four models, tighter HY OAS is the dominant driver of reduced drawdown probability. Residual upside risk at medium horizons is principally attributable to elevated equity valuations (BSEYD), early-stage labour-market softening, and still-elevated ISM Services Prices Paid.

18 Limitations and Further Research

18.1 Methodological Limitations

The present study is subject to several important limitations that affect the interpretation of results.

Look-ahead bias in indicator construction. A small number of indicators are not constructed in a fully "real-time" way. In particular, some transformations use information that would not have been available at the time—such as full-sample standardisation or rolling statistics that implicitly rely on future data. This introduces mild look-ahead contamination in the historical analysis and can make in-sample performance look slightly better than what an investor could have achieved in real time. A cleaner implementation would construct every transformed series recursively, using only information available up to each date, or using forecasted figures for the relevant indicators.

Predictor selection uses the full sample. The procedure that selects the best indicator within each thematic group is done using the full sample and then evaluated out-of-sample. A stricter evaluation would re-select the best representative inside each expanding training window, reducing the risk that the selection step benefits from hindsight. The current approach is common in screening-style empirical work but should be viewed as an upper bound on achievable real-time performance.

Class imbalance. Drawdown months are rare relative to non-drawdown months across all horizons — particularly at 1M (≈11.5% positive rate) and 3M (≈5% positive rate). Severe class imbalance affects both model training and threshold calibration: with so few positive observations, small changes in the training window can produce meaningful coefficient instability, and standard probability thresholds will mechanically generate high false-alarm rates. This sets a practical ceiling on precision that is independent of model specification.

Collinearity and coefficient instability. Financial conditions indicators (NFCI, HY OAS, VIX) are highly correlated, leading to coefficient instability within the multi-factor model. While ridge penalisation mitigates this, individual coefficient estimates should not be interpreted as precise structural parameters. The models should be evaluated collectively rather than on the significance of individual predictors.

Structural breaks. The 35-year sample spans multiple distinct monetary policy and market-structure regimes. Parameters estimated over this period are a blend of potentially distinct regimes. The first zero-lower-bound episode (2009–2015), the quantitative easing era (2009–2022), the second ZLB episode during the COVID response (2020–2022), and the current normalisation cycle may each warrant separate model parameterisations.

Unpredictable event risk. Some of the largest drawdowns are driven by shocks that macro fundamentals cannot anticipate — sudden geopolitical events, pandemics, policy surprises, or sentiment cascades. The COVID episode illustrates this clearly: fundamentals-based indicators can describe vulnerability, but they cannot reliably forecast the timing of exogenous shocks. This sets a hard ceiling on what any macro-driven drawdown prediction framework can achieve.

18.2 Further Research Directions

Several extensions to the present framework would materially improve its practical utility.

Non-linear models. The ridge logit model imposes a largely linear relationship between indicators and drawdown risk. In practice, many macro-financial effects are threshold-based — curve inversion, spreads widening past a stress level, inflation becoming persistent. Future work could test modest non-linear approaches or add interpretable threshold features and interaction terms. These approaches are most likely to improve performance at the 12M horizon, where regime-dependent dynamics are plausible and the larger event count ($n = 86$) provides sufficient data to support additional parameters. At the 3M horizon, where only 16 positive months exist, adding non-linear complexity risks overfitting and is unlikely to be productive without additional data.

International expansion. This study focuses exclusively on the US S&P 500 and US macroeconomic indicators. An international version covering Eurozone, UK, Japanese, and emerging market equity indices would test whether the same indicator groups retain predictive power globally, and would provide a much larger pool of drawdown episodes for model evaluation.

Optimal threshold selection. The 0.30 probability threshold used throughout this study is illustrative. In practice, the optimal threshold depends on the cost asymmetry between false alarms and missed events in a specific portfolio context. A formal decision-theoretic framework incorporating transaction costs, portfolio protection costs, and tail-loss exposure would produce horizon-specific and strategy-specific optimal thresholds.

Calibration testing. AUC measures discrimination, not whether predicted probabilities match realised frequencies. If the model is used for portfolio sizing (e.g., scaling risk or hedging intensity), calibration matters. Future work should test calibration to assess whether "30% risk" actually corresponds to roughly "30% event frequency" in the data.

19 Conclusion

This paper has presented a comprehensive empirical study of 30 macroeconomic and financial leading indicators for predicting S&P 500 peak-to-trough drawdowns over forecast horizons of 1, 3, 6, and 12 months. The key findings can be summarised as follows.

Firstly, market-price-based financial indicators — particularly the National Financial Conditions Index (AUC = 0.725 at 1M, 0.719 at 6M), High-Yield OAS (AUC = 0.729 at 3M, 0.663 at 6M), and VIX (AUC = 0.721 at 1M, 0.696 at 12M) — consistently outperform survey and fundamental macro indicators at all horizons, but are themselves subject to endogeneity and simultaneity concerns.

Secondly, ISM Services indicators, particularly ISM Services Employment (AUC = 0.725 at 3M), provide stronger predictive signals than their manufacturing counterparts at the critical 3-month horizon, reflecting the modern dominance of services in the US economic cycle. ISM Services Prices Paid emerges as a powerful medium and long-horizon signal (AUC = 0.633 at 6M), capturing the monetary policy tightening channel.

Thirdly, the Sahm Rule outperforms the raw unemployment rate (AUC = 0.676 vs. 0.515 at 3M), showing that rate-of-change transformations of labour market data are more informative for equity market risk than level measures at the 3M horizon, but this advantage is not robust across horizons. Conference Board Consumer Confidence, as a contrarian late-cycle signal, achieves AUC = 0.634 at 12M and carries the largest coefficient in the long-horizon model.

Fourthly, the four horizon-specific ridge logit models achieve expanding-window OOS AUCs of 0.600 (1M), 0.694 (3M), 0.737 (6M), and 0.684 (12M), all substantially exceeding the random-classifier benchmark of 0.500. The 6-month model is the strongest performer, consistent with a two-to-three quarter transmission lag from policy and credit conditions into equity drawdowns.

Fifthly, the current market assessment as of December 2025 indicates broadly benign near-term conditions (1M probability 5.0%, 3M probability 2.0%), with modestly elevated medium-term risk (6M probability 12.4%, 12M probability 8.2%). The decline from the December 2024 assessment — where 6M and 12M probabilities stood at 25.3% and 22.5% respectively — reflects the broad tightening of credit spreads and improvement in labour market conditions over 2025. HY OAS is the dominant negative driver at all horizons.

Sixthly, the historical drawdown episode analysis reveals the structural limitations of macro-based prediction for sudden exogenous events (Russia/LTCM 1998, COVID-19 2020), while demonstrating meaningful pre-crisis warning for macro-driven drawdowns (Dot-Com Bust, Great Financial Crisis, Rate Hike Drawdown 2022). The 6-month model achieves OOS recall of 0.676, correctly identifying the majority of drawdown months in the 2000–2024 out-of-sample period.

Together, these findings support a conditional reading of the predictive evidence: macro-financial indicators provide economically meaningful and statistically reliable leading signals of equity drawdown risk, particularly at the 6-month horizon, for cycles driven by credit conditions, monetary policy, and the business cycle. They are fundamentally unable to predict sudden-stop, sentiment-driven, or geopolitical shocks. Within these boundaries, the four-horizon framework presented here offers a structured, empirically grounded tool for integrating macroeconomic intelligence into risk management and tactical asset allocation decisions.

References

Baker, S. R., Bloom, N., Davis, S. J., Kost, K. J., Sammon, M. C., & Viratyosin, T. (2020). The unprecedented stock market reaction to COVID-19. *Review of Asset Pricing Studies*, 10(4), 742–758.

Bernanke, B. S., & Gertler, M. (1995). Inside the black box: The credit channel of monetary policy transmission. *Journal of Economic Perspectives*, 9(4), 27–48.

Bernanke, B. S., & Reinhart, V. R. (2004). Conducting monetary policy at very low short-term interest rates. *American Economic Review*, 94(2), 85–90.

Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic Perspectives*, 23(1), 77–100.

Clarida, R., Galí, J., & Gertler, M. (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics*, 115(1), 147–180.

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *Journal of Finance*, 46(2), 555–576.

Friedman, M., & Schwartz, A. J. (1963). *A Monetary History of the United States, 1867–1960*. Princeton University Press.

Gilchrist, S., & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4), 1692–1720.

Goyal, A., & Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455–1508.

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, 113(2), 363–398.

Jorion, P. (2000). Risk management lessons from Long-Term Capital Management. *European Financial Management*, 6(3), 277–300.

Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *Journal of Finance*, 72(3), 967–998.

Koenig, E. (2002). Using the Purchasing Managers' Index to assess the economy's strength and the likely direction of monetary policy. *Federal Reserve Bank of Dallas Economic and Financial Policy Review*, 1(6), 1–14.

Lahiri, K., & Monokroussos, G. (2013). Nowcasting US GDP: The role of ISM business surveys. *International Journal of Forecasting*, 29(4), 644–658.

Lane, P. R. (2012). The European sovereign debt crisis. *Journal of Economic Perspectives*, 26(3), 49–68.

Lleo, S., & Ziemba, W. T. (2012). Stock market crashes in 2007–2009: Were we able to predict them? *Quantitative Finance*, 12(8), 1161–1187.

Lleo, S., & Ziemba, W. T. (2015a). The bond-stock earnings yield model for stock market crash prediction: The basic idea and early applications. *Quantitative Finance Letters*, 4(1), 19–25.

Lleo, S., & Ziemba, W. T. (2015b). Some historical perspectives on the bond-stock earnings yield model for crash prediction around the world. *International Journal of Forecasting*, 31(2), 399–425.

Pástor, L., & Veronesi, P. (2006). Was there a Nasdaq bubble in the late 1990s? *Journal of Financial Economics*, 81(1), 61–100.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373–1379.

Rapach, D. E., & Zhou, G. (2013). Forecasting stock returns. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vol. 2, pp. 328–383). Elsevier.

Sahm, C. (2019). Direct stimulus payments to individuals. In H. Boushey, R. Nunn, & J. Shambaugh (Eds.), *Recession Ready: Fiscal Policies to Stabilize the American Economy* (pp. 67–92). Brookings Institution Press.

Shiller, R. J. (2000). *Irrational Exuberance*. Princeton University Press.